RELATIVISTIC EMBEDDING

submitted by

Matt James

for the degree of Doctor of Philosophy

of the

University of Bath

Department of Physics

September 2010

Summary

The growing fields of spintronics and nanotechnology have created increased interest in developing the means to manipulate the spin of electrons. One such method arises from the combination of the spin-orbit interaction and the broken inversion symmetry that arises at surfaces and interfaces, and has prompted many recent investigations on metallic surfaces.

A method by which surface states, in the absence of spin orbit effects, have been successfully investigated is the Green function embedding scheme of Inglesfield. This has been integrated into a self consistent FLAPW density functional framework based on the scalar relativistic Kölling Harmon equation. Since the spin of the electron is a direct effect of special relativity, calculations involving the spin orbit interaction are best performed using solutions of the Dirac equation. This work describes the extension of Green's function embedding to include the Dirac equation and how fully relativistic FLAPW surface electronic structure calculations are implemented.

The general procedure used in performing a surface calculation in the scalar relativistic case is closely followed. A bulk transfer matrix is defined and used to generate the complex band structure and an embedding potential. This embedding potential is then used to produce a self consistent surface potential, leading to a Green's function from which surface state dispersions and splittings are calculated. The bulk embedding potential can also be employed in defining channel functions and these provide a natural framework in which to explore transport properties. A relativistic version of a well known expression for the ballistic conductance across a device is derived in this context. Differences between the relativistic and non-relativistic methods are discussed in detail.

To test the validity of the scheme, a fully relativistic calculation of the extensively studied spin orbit split L-gap surface state on Au(111) is performed, which agrees well with experiment and previous calculations. Contributions to the splitting from different angular momentum channels are also provided. The main advantages of the relativistic embedding method are the full inclusion of the spin orbit interaction to all orders, the true semi infinite nature of the technique, allowing the full complex bands of the bulk crystal to be represented and the fact that a only small number of surface layers is needed in comparison to other existing methods.

PUBLICATIONS

The key developments and results in this thesis have been presented in the following article [1]

M. James & S. Crampin, Relativistic Embedding Method: The Transfer Matrix, Complex Band Structures, Transport and Surface Calculations. *Phys. Rev. B*, **81**(15), 155439 (2010).

Contents

Chapter 1

INTRODUCTION

All of our familiar world is constructed of atoms, themselves made from the protons and neutrons of the nucleus, and electrons, the 'glue' which binds nuclei together into solids, liquids and molecules. The arrangement of the atoms of a solid into a particular crystal structure is due to the interaction of the electrons within it and electron excitations determine its optical, electrical and magnetic properties. The shape and action of the proteins and molecules which govern the functioning of biological systems, including ourselves, are determined by the behaviour of the electrons which they contain and an understanding of this behaviour is essential to the development of drugs and treatments for disease. The increasing reliance of the modern world on computers demands that we are able to manipulate extremely accurately the movement of electrons through devices on an ever decreasing scale. Consequently, one of the foremost challenges of theoretical physics is to develop methods and models which enable us to completely understand the fundamental behaviour of collections of interacting electrons.

In the 1970s, features on an integrated circuit were in the region of 10 μm in size. In 2010, length scales on commercially mass produced chips are roughly 60 nm and may reach or indeed surpass 30 nm within the next five years. This decrease in scale cannot continue indefinitely as there is a limit to size beyond which conventional devices simply will not work. In very small semiconductor transistors the variation in distribution of dopants from device to device may lead to dramatic differences in performance and the tiny size of the insulator gate allows electrons to tunnel across. The rapid approach of this practical size limit on current technologies has given rise to the field of nanoelectronics and spintronics. Relevant major breakthroughs in this area have been the discovery of Giant Magnetoresistance (GMR) and Tunneling Magnetoresistance (TMR) in magnetic multilayers, leading to the field of spintronics, and the resulting investigation of electron spin transport through nanodevices. Development of devices employing these effects requires a detailed knowledge of the consequences of the spin orbit interaction at material surfaces and interfaces.

The arrival of quantum mechanics in the early part of the 20th century was a

vast leap forward in understanding the behaviour of electrons and other fundamental particles. Many simple systems, such as the hydrogen atom and harmonic oscillator, which could be investigated analytically, were quickly solved but more complex problems which did not permit exact solutions proved elusive. Even a small sample of material, a few grammes of an elemental solid, contains of the order of 10^{23} electrons, which not only interact with the atomic nuclei but with each other, and it was not until the advent of the computer that the study of such systems could be realistically attempted. Still, the single particle picture of quantum mechanics needed adaptation to allow practical investigation of collections of multiple interacting particles to proceed. Density Functional Theory (DFT) has become a standard framework within which many such calculations are performed yet the continuing demand for ever smaller and more complex electronic devices requires more accurate and varied theoretical models which are necessarily more computationally demanding, requiring continual refinement and development of existing techniques.

Calculation of ground state properties of bulk systems within DFT is simplified by exploiting the periodicity of the crystal lattice, but the presence of a surface destroys translational symmetry and allows electrons to become trapped in the surface region. These surface states decay exponentially into the bulk crystal and are of extreme significance to transport properties due to their influence on electron scattering at the surface. A common technique used to calculate the surface density of states, thus determining the dispersion of surface states, is to use a supercell or slab method. Here, three dimensional translational symmetry is restored by introducing a repeating cell containing several layers of atoms and a portion of vacuum to represent the surface, and imposing periodic boundary conditions. The cell must contain enough atoms that those in the layer farthest from the surface are bulk–like and enough vacuum that surface states in adjoining cells do not interact. There are two main disadvantages to the supercell approach. Firstly, in order to reproduce bulk like behaviour and non–interacting surface states it may be necessary to include 20 or more layers in the cell and several atomic units (a.u.) of vacuum, leading to calculations which can be computationally very demanding. Secondly, the bulk band structure is not faithfully reproduced since the system is not truly semi–infinite, more resembling a thin film, resulting in a series of discrete states rather than a true continuum. This can make it hard to separate out surface states from the bulk states or distinguish surface resonances from true surface states.

A method which is extremely well suited to calculations involving semi–infinite systems is the embedding method of Inglesfield [2]. Using this approach, a Green function is solved for explicitly in typically only 2 or 3 layers of atoms in the surface region and a small portion of vacuum. The bulk and vacuum continua are represented by an energy dependent *embedding potential*, essentially a surface logarithmic derivative, the embedding potential itself being calculated from a surface expansion

of the bulk Green function. This method reproduces exactly the bulk band structure and alongside surface electronic structure calculations has been adapted to perform investigations of ballistic transport in layered systems and molecules attached to semi–infinite leads.

A good description of surfaces is provided by the embedding formulation and a way of using this to determine the rôle of the spin orbit interaction in electronic structure and transport properties would provide valuable insight into the future of spintronic devices. Although spin orbit effects may be incorporated into the Schrödinger equation on an *ad hoc* basis to a first order approximation with some satisfactory results, electron spin and its consequences are truly a relativistic phenomenon. The Koelling–Harmon scheme has been used in the embedding framework but this contains only the non spin related relativistic effects. A complete description is therefore only attainable by starting from the Dirac equation. Previously, the basic formulation of embedding has been extended to the Dirac equation [3] and this thesis expands upon this work and develops it to a point where practical calculations of dispersion relations of surface states can be calculated. In addition, a starting point from which transport calculations may be performed is introduced.

In the remainder of this chapter we discuss the motivation for including relativistic effects in electronic structure calculations and show that the most important of these effects is the spin-orbit interaction, and argue that there is a definite class of systems where this is of crucial significance in the development of new technologies. Before introducing the embedding method, we present, in chapter 2, a detailed discussion of the Dirac equation, concentrating on the solutions and properties which are most relevant to the relativistic embedding method. We also consider the differences between the Dirac equation and non-relativistic quantum mechanics, focussing in particular on the fact that electron spin, and hence the spin-orbit interaction, is a natural consequence of the relativistic theory, providing further justification for a relativistic description of electronic structure. In this section we also introduce Green's functions, which are a powerful tool in electronic structure calculations and central to the embedding method. We also compare the relativistic and non-relativistic Green's functions. In chapter 3, we present a brief description of Density Functional Theory and demonstrate how this is formulated and applied to practical calculations, again discussing the non-relativistic and relativistic cases.

Having presented the motivation and necessary background in the previous chapters, in chapter 4 we (at last!) introduce the Green's function embedding method. This chapter is divided into three main parts. The first describes, in detail, the non-relativistic embedding method for the general case of a system divided into two distinct regions, and then demonstrates an application to a simple, model system to illustrate how it works in practice. The second part provides a similar description of the relativistic extension of the embedding method. This is done in the same way

as the non-relativistic case to allow comparison of the two. The final part of the chapter is devoted to a brief summary of how the non-relativistic method has been applied to realistic problems and provides a background for the following chapters.

The rest of the thesis is a detailed description of how the relativistic embedding method can be used in practice, concentrating especially on surface electronic structure, but also providing a useful framework for transport studies. Chapter 5 introduces the idea of the transfer matrix, which is a particularly useful way of generating the embedding potential that describes the substrate in a surface calculation. The transfer matrix also provides the bulk complex band structure and links the embedding method with transport calculations via the idea of channel functions. We use it to derive the relativistic version of a well known formula for the conductance through a device which could well provide a useful starting point for further investigations along these lines.

The final chapter is concerned with implementation of the embedding method for the investigation of surface electronic structure and, following a presentation of the method used for such calculations, contains the results of a test calculation on the much studied Au(111) surface state, which exhibits a strong spin-orbit splitting. We show that the method reproduces the results of experiment and previous theoretical studies, over which it has particular advantages. We also present some novel results which give new insight into the mechanism of the spin-orbit splitting of surface states and which provides a strong motivation for this relativistic implementation of the embedding method in further studies.

We will begin with a question...

1.1 Is Relativity Important?

Relativity is concerned, as its name suggests, with physical phenomena viewed by observers in relative motion to one another. Physical laws are expressed relative to space, (x, y, z), and time, t, coordinates which are stationary relative to a given observer. These coordinates are the observer's *reference frame* and relativity tells us how these coordinate systems are related between frames. Special relativity considers reference frames moving with a constant velocity relative to each other, called *inertial* reference frames. If we have an observer O at position (x, y, z) and time t and a second observer O' at (x', y', z') and time t' moving along the x-axis at speed v relative to O, then the non-relativistic Galilean transformations relate the coordinate systems of O and O'

$$x' = x - vt, \qquad y' = y, \qquad z' = z, \qquad t' = t. \tag{1.1}$$

The principle of relativity says that *all physical laws must be the same for all observers*, or in other words, any equation which describes a physical process must be

the same (invariant) when we move from (x, y, z, t) to (x', y', z', t'). The Schrödinger equation is invariant under a Galilean transformation, as are Newton's laws of motion.

Several observations and experiments in the late nineteenth and early twentieth century led Einstein to reformulate the principle of relativity by adding a second part; *The speed of light, c, is a maximum speed and is always measured to be the same value in all inertial reference frames, irrespective of the relative velocity of observers.* This leads to a new set of coordinate transformations, the *Lorentz transformations* [4]:

$$x' = \gamma\left(x - vt\right), \qquad y' = y, \qquad z' = z, \qquad t' = \gamma\left(t - \frac{xv}{c^2}\right), \qquad \gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}. \tag{1.2}$$

The Lorentz transformations lead to a range of counter intuitive effects: A moving object appears shorter in the direction of its motion when measured by a stationary observer (Lorentz contraction). If we measure the mass of an object when it is at rest, in its rest frame, and then again when it is moving relative to us, we see that its mass appears to increase. Two events which appear to be simultaneous in one inertial frame may not appear simultaneous in another. The energy, E, of a free particle of mass m travelling at velocity v is given by

$$E = \gamma m c^2 \tag{1.3}$$

and if we expand γ, from (1.2), in powers of v^2/c^2 we arrive at the most famous effect of special relativity: the appearance of a particle's *rest energy*, $E_0 = mc^2$ in the expansion of its kinetic energy,

$$E = mc^2 + \frac{1}{2}mv^2 - \frac{1}{8}\frac{mv^4}{c^2} + \cdots . \tag{1.4}$$

This expansion also highlights an important point; if $v << c$ we recover the classical energy $mv^2/2$ (plus the rest mass) and for low velocities, the effects of relativity are small. In fact, the Lorentz transformations reduce to the Galilean transformations for $v << c$. We would therefore expect relativistic effects to be significant only when v is an appreciable fraction of c or, equivalently, at high energies.

The Lorentz and Galilean transformations in (1.1) and (1.2) and the kinetic energy expansion can be used to illustrate another useful idea; that of *non-relativistic limits*. The relativistic corrections to the energy are in powers[‡] of c and if we let $c \to \infty$ we obtain the non-relativistic energy, as long as we ignore the rest mass. This idea of letting $c \to \infty$ is consistent with non-relativistic mechanics since there is no upper limit to velocities. When we have a relativistic expression we can see

[‡]This leads to the rather confusing convention that a first order relativistic correction is actually proportional to $1/c^2$

how it compares to the non-relativistic expression by expanding in powers of c and letting $c \to \infty$. In general then, if we have terms in a relativistic expression which depend on powers of $1/c$ then they are relativistic corrections, while terms which are proportional to powers of c, such as the rest energy, are effects which have no non-relativistic equivalent.

In special relativity, any equation describing some physical process must be invariant under the Lorentz transformations (1.2) and they may be regarded as a set of symmetries satisfied by time and space. As a result they may be considered as a fundamental starting point and leads to a deep understanding of physical laws and processes (see for example [5]). The Lorentz transformations 'mix' space and time coordinates; motion causes a rotation of the space and time axes when transforming between inertial frames. This leads to the idea of enhanced symmetries shared by space and time, joining them into a single concept, spacetime, in contrast to the non-relativistic notion that they are separate.

The relativistic corrections we have discussed so far only become significant when the particle's speed is an appreciable fraction of c. A test of how 'relativistic' a particle is can therefore be performed by comparing its velocity with c, or its kinetic energy with its rest mass. From the simple Bohr model , the velocity of an electron in the n^{th} energy level of a Hydrogenic atom of atomic number Z is $v_n = 2.1877 \times 10^6 Z/n\, ms^{-1}$. For the $n = 1$ level of Hydrogen this gives $v/c = 0.007$ and we would not expect significant relativistic effects due to the electron velocity. However, for gold, $Z = 79$ and $v/c = 0.576$, thus relativity should play an appreciable part [†]. These relativistic changes in velocity (and hence mass) in heavier elements can indeed significantly affect their physical properties. For example, the $1s$ electrons become more tightly bound as a consequence of relativistic effects and orbit closer to the nucleus. The other s electrons orbitals also contract to maintain their orthogonality to the $1s$ state through oscillations in the radial part of the wavefunction. This creates more screening of the nuclear charge and the outer d and f orbitals expand. The raising of the $5d$ and lowering of the $6s$ energies in Au account for the difference in ionisation energies between Au and the elements above it in the periodic table: 9.22 eV for gold, 7.76 eV for silver and 7.72 eV for copper.

There is however another consequence of relativity which, particularly in the context of this work, is more significant than mass-velocity effects; electron spin.

[†]It is interesting to note that for $Z = 137$, $v/c \approx 1$, yet this simple model does not take into account important effects such as the non–point charge nature of the nucleus and small orbital radius of the inner electrons leading to high electric fields. These fields cause breakdown of the vacuum and generate positrons which screen the nuclear charge, producing effective atomic numbers less than the nuclear Z. An estimate for maximum Z is 173 [6].

1.1.1 Non-Relativistic Spin

Maxwell's equations of electromagnetism are invariant with respect to Lorentz transformations and are consistent with special relativity[‡]. In a similar way to the concept of spacetime, these symmetries lead to the union of the electric and magnetic fields associated with charged particles into the electromagnetic field and explains the association of a magnetic field with a moving charge. Due to its motion around the nucleus, an electron in an atom has a magnetic moment associated with its orbital angular momentum, which is quantised. However, an electron also has a magnetic moment in its rest frame, its *spin*, which is also quantised and in analogy with the orbital moment the spin can also be associated with an angular momentum, hence its name. We will see later on that the spin of the electron is implicit in the Dirac equation, which is the Lorentz invariant version of the Schrödinger equation in quantum theory. This suggests that spin is a relativistic phenomenon and is a result of the symmetries of spacetime. The effects of spin are therefore relativistic effects. Before discussing these effects in detail it is worth looking at how spin fits into non-relativistic quantum mechanics.

The orbital angular momentum operator, $\hat{\mathbf{L}}$ in non-relativistic quantum mechanics is defined by

$$\hat{\mathbf{L}} = \hat{\mathbf{r}} \times \hat{\mathbf{p}} = -\mathrm{i}\hbar \mathbf{r} \times \nabla \tag{1.5}$$

and satisfies a set of commutation relations [8]

$$[L_i, L_j] = \mathrm{i}\hbar\epsilon_{ijk}L_k, \qquad \left[\mathbf{L}^2, L_i\right] = 0, i = x, y, x, \text{ etc} \tag{1.6}$$

where we have dropped the 'hats' on the operators, and ϵ_{ijk} is the *Levi-Civita* pseudotensor:

$$\epsilon_{ijk} = \begin{cases} 1 & \text{for cyclic permutations of} \quad x, y, z \\ -1 & \text{for anti-cyclic permutations of} \quad x, y, z \\ 0 & \text{otherwise} \end{cases} . \tag{1.7}$$

Although we can think of the electron spin as an angular momentum, there is a problem with this definition. Electrons in non-relativistic quantum mechanics are point particles and so the definition of angular momentum in (1.5) at $\mathbf{r} = 0$ is problematic. It is easier to use the commutation relations (1.6) as the fundamental definition of angular momentum and to think of the spin as a degree of freedom which is described by an operator, $\mathbf{S}$, which satisfies the angular momentum commutation relations

$$[S_i, S_j] = \mathrm{i}\hbar\epsilon_{ijk}S_k, \qquad \left[\mathbf{L}^2, S_i\right] = 0. \tag{1.8}$$

The orbital angular momentum operators have eigenvalues, l, and eigenfunctions

[‡]In fact, Lorentz discovered his transformation equations by studying Maxwell's equations [7].

$\psi_l^{m_l}$

$$\begin{aligned} \mathbf{L}^2\psi_l^{m_l} &= \hbar^2 l(l+1)\psi_l^{m_l} \qquad l = 0, 1, 2, 3, \cdots \\ L_z\psi_l^{m_l} &= \hbar m_l \psi_l^{m_l} \qquad -l \leq m_l \leq l \end{aligned} \tag{1.9}$$

where l is the azimuthal orbital angular momentum quantum number and m_l is the magnetic orbital angular momentum quantum number. We can define similar eigenvalues and eigenfunctions for the spin

$$\mathbf{S}^2\chi_{1/2}^{m_s} = \hbar^2 s(s+1)\chi_s^{m_s}, \qquad S_z\chi_s^{m_s} = \hbar m_s \chi_s^{m_s}, \qquad -s \leq m_s \leq s. \tag{1.10}$$

From experiment[†] it is known that there are two possible values of m_s for the electron, and $m_s = \pm 1/2$ for $s = 1/2$. When $m_s = +1/2$, the z-component of the spin is parallel to the z-axis, 'spin up' and for $m_s = -1/2$ it is antiparallel, 'spin down'.

Spin does not arise naturally from the Schrödinger equation and we must put it in 'by hand'. To do this we define the eigenfunctions (usually called 'spinors') in spin space

$$\phi_\uparrow = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{for} \quad m_s = +\frac{1}{2}, \qquad \phi_\downarrow = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{for} \quad m_s = -\frac{1}{2}. \tag{1.11}$$

With these definitions, the spin operator can be written in terms of the *Pauli spin matrices*

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad S_x = \frac{\hbar}{2}\sigma_x \quad \text{etc.} \tag{1.12}$$

The Pauli matrices satisfy the same commutation relations as the spin operator. The σ_z matrix is diagonal because the spin is quantised along the z-axis.

The Schrödinger equation describes a system where the external potential does not describe an interaction which couples the spins, and we can write the wavefunction as a product of spatial and spin parts

$$\Psi(\mathbf{r}) = \sum_{m_s} \psi(\mathbf{r})\phi_{m_s} \tag{1.13}$$

and we have a separate equation for each spin. This is only true if the spin operator commutes with the potential.

The spin magnetic moment of the electron, $\boldsymbol{\mu}_s$, is written

$$\boldsymbol{\mu}_s = g_s \mu_B \mathbf{S} \tag{1.14}$$

[†]These are Stern-Gerlach experiments where a beam of electrons is passed between magnets and splits into two separate beams corresponding to the different values of the spin magnetic quantum number [8]

in terms of the electron spin g-factor, g_s, and the Bohr magneton, $\mu_B = e\hbar/2m$ [9]. If there is an external magnetic field, $\mathbf{B}$ then the spin magnetic moment will interact with the field to give a potential energy

$$-\boldsymbol{\mu}_s \cdot \mathbf{B} = \frac{e\hbar}{2m}\boldsymbol{\sigma} \cdot \mathbf{B} \tag{1.15}$$

where we have used $g_s = 2$[‡] for the spin g-factor [8]. Writing the magnetic field in terms of a vector potential, $\mathbf{B}(\mathbf{r}) = \nabla \times \mathbf{A}(\mathbf{r})$ gives the Pauli equation describing a spin-1/2 particle in a magnetic field

$$\frac{1}{2m}\left[-\mathrm{i}\hbar\nabla - e\mathbf{A}(\mathbf{r})\right]^2 \Psi(\mathbf{r},t) - \frac{e\hbar}{2m}\boldsymbol{\sigma}\cdot\mathbf{B}(\mathbf{r})\Psi(\mathbf{r},t) + V(\mathbf{r},t)\Psi(\mathbf{r},t) = \mathrm{i}\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r},t) \tag{1.16}$$

where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$.

As well as interacting with an external magnetic field, the spin magnetic moment can interact with the field generated by the orbital motion of the electron; *spin-orbit coupling*. It is possible to derive a potential describing spin-orbit coupling in a spherically symmetric potential, $V(r)$, from classical electromagnetism and the Lorentz transformations [10]. The potential is

$$H_{\mathrm{SO}} = -\frac{1}{2m^2c^2}\frac{1}{r}\frac{\mathrm{d}V(r)}{\mathrm{d}r}\mathbf{L}\cdot\mathbf{S}. \tag{1.17}$$

If we put this into the Schrödinger equation (for now neglecting the time dependence and any magnetic fields) we have

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(r)\right]\Psi(\mathbf{r}) - \frac{1}{2m^2c^2}\frac{1}{r}\frac{\mathrm{d}V(r)}{\mathrm{d}r}\mathbf{L}\cdot\mathbf{S}\Psi(\mathbf{r}) = H_0\Psi(\mathbf{r}) + H_{so}\Psi(\mathbf{r}) = E\Psi(\mathbf{r}). \tag{1.18}$$

where $V(r)$ is just a scalar, spin-independent potential. The operators $\mathbf{L}^2$, L_z, $\mathbf{S}^2$ and S_z commute with H_0 but L_z and S_z do not commute with H_{so}. This means that eigenfunctions of S_z and L_z cannot also be eigenfunctions of $H_0 + H_{so}$. One can, however, define a total angular momentum operator, $\mathbf{J} = \mathbf{L} + \mathbf{S}$ so that $H_0 + H_{so}$ commutes with $\mathbf{J}^2$ and J_z. The eigenvalues of $\mathbf{J}^2$ are $\hbar^2 j(j+1)$ where $j = l \pm s$ and $-j \leq m_j \leq j$. Without the spin orbit interaction, the spin up and spin down eigenstates for a given l have the same energy (degenerate) but the spin orbit interaction breaks the symmetry between spin up and down and the degeneracy is lifted, giving rise to a splitting in energy for the two states. This is the origin of the fine structure in the line spectra of atoms. When we discussed the effects of the velocity/mass corrections in atoms we saw that they would be negligible for

[‡]Quantum electrodynamic corrections to the g-factor due to the interaction of the electron with its own magnetic field predict a value of 2.002319 which agrees with the experimental value to well within the limits of accuracy. In fact calculation and measurement of the g-factor with increasing precision continue to confirm the astonishing accuracy of quantum electrodynamics.

Hydrogen. However, the fine structure in Hydrogen is measurable and since the spin-orbit interaction is relativistic in origin (it has a $1/c^2$ factor), relativistic effects are observed even in lighter atoms.

The dependence of the spin-orbit term in (1.18) on $1/c^2$ means that it is in general a small correction to the Hamiltonian, since the speed of light is large, $c = 299792458\, ms^{-1}$. Energy corrections due to the spin-orbit interaction in non-relativistic quantum mechanics may be treated within perturbation theory. The first order energy correction in for the n^{th} energy level in a Hydrogenic atom of atomic number Z is [8]

$$E_{\mathrm{SO}}^{(1)} = \frac{Z^4 m e^8}{2(4\pi\varepsilon_0\hbar)^4 c^2} \frac{j(j+1) - l(l+1) - \frac{3}{4}}{n^3 l(l+1)(2l+1)} \tag{1.19}$$

The prefactor is $\approx 6.6 \times 10^{-6} Z^4\, eV$ so the splittings in hydrogen are tiny (yet still observable) but the dependence on Z^4 means that we already have splittings of around $1\, eV$ for Calcium at $Z = 20$. In reality the Z^4 dependence breaks down for all but the simplest elements due to the screening of the nuclear charge by the extra electrons. Nevertheless the spin orbit splittings are still significant for heavier elements.

1.1.2 The Spin-Orbit Effect at Surfaces

As well as giving rise to the fine-structure of atoms, the spin-orbit interaction has significant consequences at the surfaces of solids. Electrons in a bulk material can be labelled [5] by their wavevector $\mathbf{k}$ and spin s, and in the absence of magnetic fields, time reversal symmetry means that states with opposite spin and momentum are degenerate, $E(\mathbf{k}, \uparrow) = E(-\mathbf{k}, \downarrow)$, known as Kramer's degeneracy [5]. If the crystal structure of the material has inversion symmetry, i.e. looks the same if all atoms are reflected through a particular point $\mathbf{r} \to -\mathbf{r}$, then $E(\mathbf{k}, s) = E(-\mathbf{k}, s)$ [5]. If both of these symmetries exist then $E(\mathbf{k}, \uparrow) = E(\mathbf{k}, \downarrow)$ and both spin states are degenerate. For example, in Gold, for which $Z = 79$, there is no spin splitting of the bulk bands despite the relatively high strength of the spin orbit interaction due to the large atomic number. However, inversion symmetry is broken at the surface of a solid, and the two spin states are no longer forced to be degenerate, $E(\mathbf{k}, \uparrow) \neq E(\mathbf{k}, \downarrow)$ indicating a lifting of the spin degeneracy, which has been observed experimentally in surface states [11].

In a bulk material electron states are delocalised throughout the crystal, giving rise to the energy bands. At the surface however, it is possible for electrons to become 'trapped' between the surface barrier and a bulk band gap, forming states which are localised in the direction normal to the surface.

In the region close to a surface, the potential differs from that in the bulk, and the electrons feel a 'surface' potential, $V_z(\mathbf{r})$ and we can associate an electric field

with this potential, $\mathbf{E}(\mathbf{r}) = -\nabla V_z(\mathbf{r})$. An electron moving with velocity $\mathbf{v}$ in an electric field $\mathbf{E}$ defined in some global reference frame will experience a magnetic field $\mathbf{B}$ in its rest frame

$$\mathbf{B} = \frac{1}{c^2}\mathbf{v} \times \mathbf{E} = -\frac{1}{c^2}\mathbf{v} \times \nabla V_z(\mathbf{r}) = \frac{1}{c^2}\nabla V(_z\mathbf{r}) \times \mathbf{v} \tag{1.20}$$

via the Lorentz transformations [12]. The spin of the electron will couple to this magnetic field through its magnetic moment,

$$-\boldsymbol{\mu} \cdot \mathbf{B} = -\frac{e\hbar}{2m}\boldsymbol{\sigma} \cdot (\nabla V_z(\mathbf{r}) \times \mathbf{v}) = \frac{e\hbar}{2m}\nabla V_z(\mathbf{r}) \cdot (\boldsymbol{\sigma} \times \mathbf{v}). \tag{1.21}$$

We can substitute in for the velocity, $\mathbf{v} = \mathbf{p}/m$ to give a spin-orbit interaction for electrons at the surface

$$H_{SO} = \frac{e\hbar^2}{2m^2c^2}\nabla V(\mathbf{r}) \cdot (\boldsymbol{\sigma} \times \mathbf{p}). \tag{1.22}$$

For transparency, we now consider the simple case of a two-dimensional electron gas (2DEG), such as that found in semiconductor heterostructures, where electrons are confined in a small region near the surface and are free to move in the surface plane. If the surface is in the x-y plane, then the electrons can be described by Bloch states with $\mathbf{k}_\parallel = (k_x, k_y)$. The Hamiltonian in this case is the *Rashba Hamiltonian* found in the investigation of semiconductor heterostructures and quantum wells [13],

$$H = -\frac{\hbar^2}{2m^*}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + \alpha_R(\sigma_x p_y - \sigma_y p_x). \tag{1.23}$$

The effective mass of electrons in the valence and conduction bands is m^* and the *Rashba parameter*, α_R, is system specific with a magnitude proportional to the structural asymmetry described by $\nabla V_z(\mathbf{r})$ [14]. The part of the wavefunction for the electrons with momentum parallel to the surface can be written as a product of spatial and spin parts [15]

$$\psi(\mathbf{k}_\parallel) = \mathrm{e}^{\mathrm{i}\mathbf{k}_\parallel \cdot \mathbf{r}_\parallel}\chi \tag{1.24}$$

and the Hamiltonian can be written in the matrix form

$$H = \begin{bmatrix} \dfrac{\hbar^2}{2m}|\mathbf{k}_\parallel|^2 & \alpha_R(\mathrm{i}k_x + k_y) \\ -\alpha_R(\mathrm{i}k_x - k_y) & \dfrac{\hbar^2}{2m}|\mathbf{k}_\parallel|^2 \end{bmatrix}. \tag{1.25}$$

The eigenvalues, E, can be calculated directly from the determinant $|H - E| = 0$ are given by

$$E_\pm = \frac{\hbar^2}{2m}|\mathbf{k}_\parallel|^2 \pm \alpha_R|\mathbf{k}_\parallel|. \tag{1.26}$$

We can write

$$k_x \pm \mathrm{i}k_y = |\mathbf{k}_\parallel|\mathrm{e}^{\mathrm{i}\theta}, \qquad k_x = \cos\theta,\ k_y = \sin\theta, \qquad \chi = \begin{pmatrix} a \\ b \end{pmatrix}, \tag{1.27}$$

and the normalised spin part of the wavefunction, χ, is given by solving

$$\begin{bmatrix} \mp 1 & \mathrm{i}\mathrm{e}^{-\mathrm{i}\theta} \\ -\mathrm{i}\mathrm{e}^{\mathrm{i}\theta} & \mp 1 \end{bmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 0, \quad \text{with} \quad a^2 + b^2 = 1. \tag{1.28}$$

The eigenstates of the Rashba Hamiltonian are therefore

$$\psi_\pm(\mathbf{k}_\parallel) = \frac{\mathrm{e}^{\mathrm{i}\mathbf{k}_\parallel \cdot \mathbf{r}_\parallel}}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm\mathrm{i}\mathrm{e}^{\mathrm{i}\theta} \end{pmatrix}. \tag{1.29}$$

These can be used to calculate the spin polarisation, $\mathbf{P}_\pm(\mathbf{k}_\parallel)$ from

$$\mathbf{P}_\pm(\mathbf{k}_\parallel) = \psi_\pm^\dagger(\mathbf{k}_\parallel)\boldsymbol{\sigma}\psi_\pm(\mathbf{k}_\parallel) \tag{1.30}$$

which is given by

$$P_\pm^x(\mathbf{k}_\parallel) = \mp\frac{k_y}{|\mathbf{k}_\parallel|}, \qquad P_\pm^y(\mathbf{k}_\parallel) = \pm\frac{k_x}{|\mathbf{k}_\parallel|}, \qquad P_\pm^z(\mathbf{k}_\parallel) = 0. \tag{1.31}$$

This shows that for electron states which are confined normal to the surface and delocalised in the plane parallel to the surface give rise to a dispersion with a spin-splitting which is linear in $|\mathbf{k}_\parallel|$ and the magnitude of the splitting can be characterised by a Rashba parameter, α_R. The spin of the electrons is polarised in the surface plane and the polarisation rotates clockwise for the state whose energy is shifted by $-\alpha_R$ (the 'inner' state) and anticlockwise for the state shifted by $+\alpha_R$ (the 'outer' state). The dispersion curve of these states is shown in fig 1-1 along with the spin polarisation in the two-dimensional Brillouin zone (k_x, k_y).

The simple model applied here to the spin-splitting of surface states predicts a very small Rashba type splitting of surface states, of the order of 10^{-6} eV [11]. These splittings are too small to be observed by Angle-Resolved-Photoemission Spectroscopy (ARPES), yet ARPES experiments on the Au(111) surface reveal a splitting of 110 meV at the Fermi energy (see figure 1-2) [11], [15], [16], [17]. The reason for this is a combination of the Rashba effect due to the inversion symmetry breaking at the surface and the spin-orbit interaction of the atoms in the crystal [18].

1.2 Spintronics

The basic idea of spintronic devices is that the electric current can be controlled by the manipulation of the spin of the electrons which carry the current through

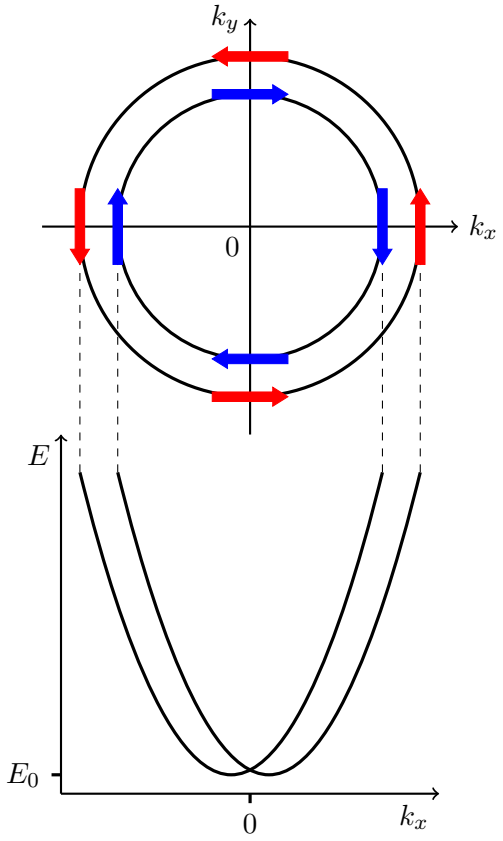

Figure 1-1: Schematic representation of equations 1.26 and (1.31), the Rashba splitting of parabolic surface state bands and the spin polarisation. The bottom part of the figure shows the surface state dispersion along k_x and the top part the spin-polarisation of the two energy states in the surface Brillouin zone.

the device. The interest in such devices began with the independent discovery of Giant Magnetoresistance (GMR) by Fert and Grünberg in the 1980s [19], [20]. They found that an electrical current passing through ferromagnetic films separated by non-magnetic spacer layers has a resistivity which depends on the change in the relative alignment of the magnetisation in the magnetic layers from ferromagnetic to antiferromagnetic. This makes it possible to turn the binary information contained in a two state magnetic configuration into an electric current, an effect which is already exploited in commercially available hard disk read heads. The realisation that the spin of the electron could be employed in electrical devices has generated a vast amount of research into potential applications [21], [22].

In the thin film magnetic structures described above, the tuning of the spin-current is achieved by switching the orientation of the magnetic field in the magnetic layers. In practice, it would be more desirable to be able to achieve this switching with an electric field in much the same way as in, for example, conventional transistors. In semiconductor devices it is much easier to generate electric fields than it is to integrate externally controllable magnetic fields. Datta and Das [13] proposed

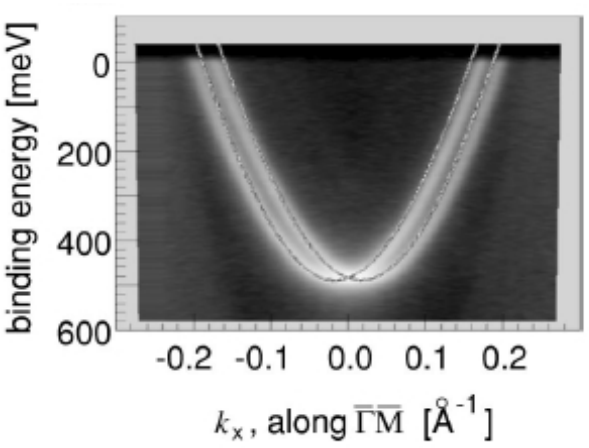

Figure 1-2: Results of an ARPES measurement of the Au(111) surface state. The dispersion is parabolic and shows a clear energy splitting which is linear in $|\mathbf{k}_{||}|$, characteristic of a Rashba type spin-orbit split Shockley surface state (compare with figure 1-1). The results are reproduced from ref. [16].

the idea of a spin field effect transistor. The source and drain are ferromagnetic materials with parallel magnetic alignments from which spin polarised electrons can be injected into a region containing a two-dimensional electron gas(2DEG), such as a InGaAs/InAlAs heterojunction (see figure 1-3). As the spin-polarised electrons travel through the 2DEG, their spins precess due to the spin orbit interaction and the material properties of the channel, the Rashba effect. The gate voltage allows the tuning of the magnitude of the precession. If the precession period is much longer than the time taken for the electrons to travel from source to drain, then they will leave the channel with their spins pointing in the same direction as when they entered, and the current will be large. If the spins reverse as they travel across the channel, then the current will be small.

The discovery of the Rashba effect at metallic surfaces is interesting since it leads to the possibility of non magnetic selection of spin polarised electrons. The surface state energy splitting means that application of a bias voltage could potentially be employed to allow conduction of the separate spin states. The ability to engineer new materials with large Rashba splittings is therefore necessary for the development of spintronic devices and an understanding of the underlying mechanisms of the effect in existing systems is crucial. Since the effect was discovered on the Au(111) surface, this was initially the most frequently studied system. However there has been recent theoretical and experimental work on a number of different metallic and semiconductor surfaces which also exhibit Rashba type spin orbit split surface states [18], [23], [24]. As we have seen, for an archetypal Rashba split surface state, the spin polarisation vector lies in the surface plane. However, recent work has shown that for more complicated systems such as semimetal overlayers on noble metal surfaces and ultrathin metallic films on semiconductor substrates, it may be

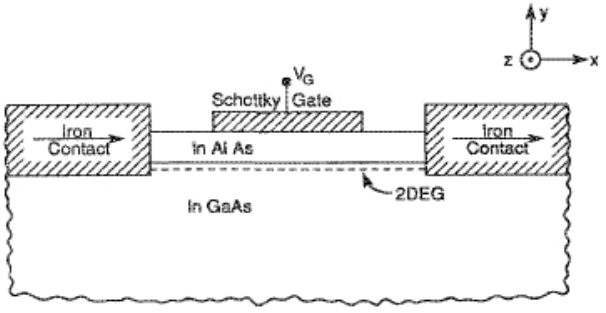

Figure 1-3: Schematic view of a spin field effect transistor (reproduced from ref [13]). Spin polarised electrons are injected from the ferromagnetic source on the left, travel across of two dimensional electron gas (2DEG) and exit at the ferromagnetic drain on the right. The gate electrode allows the spin of the electrons to be switched as they travel through the channel, via the spin orbit interaction, and the spin current flowing from source to drain can be modulated.

possible to engineer the spin polarisation to lie out of the surface plane [25], [26].

It is clear, then, that an accurate framework by which the Rashba effect at metallic surfaces and can be studied theoretically is therefore extremely important, and since the effect is due to the spin-orbit interaction, any such framework must therefore include all of the spin-orbit effects. The embedding method is ideal for treating surfaces and interfaces and its relativistic extension will implicitly include the spin-orbit interaction to all orders, enabling suitably accurate modelling of Rashba-split surfaces states and spin transport. A particular motivation for a relativistic embedding method is therefore to provide a good understanding of spin-orbit effects at surfaces and interfaces and their spin-transport properties, which is crucial in the development of practical spintronic devices in the future.

Chapter 2

RELATIVISTIC QUANTUM MECHANICS

The aim of this chapter is to introduce some of the basic concepts and results from the Dirac theory of quantum mechanics that will be needed for the subsequent development of the relativistic embedding method. The first part introduces the Dirac equation and the motivation for it. The second part presents plane wave and atomic solutions of the Dirac equation which will be called upon later. These solutions highlight some the differences and difficulties of the Dirac theory when compared with a non-relativistic treatment. The idea of non-relativistic limits is introduced to enable this comparison. Particular attention is paid to the origins and consequences of spin in the Dirac theory. Finally, the Dirac Green's function is introduced and some properties and representations relevant to later chapters are discussed.

2.1 The Dirac Equation

Non-relativistic quantum mechanics is based on the Schrödinger equation. This can be derived by writing the total energy, E, of a particle as a sum of kinetic, T, and potential energy terms, V, using the non-relativistic expression for the kinetic energy in terms of the momentum, $\mathbf{p}$ and mass, m of the particle:

$$E = T + V = \frac{p^2}{2m} + V(\mathbf{r}). \tag{2.1}$$

The momentum and energy are then replaced by operators acting on the wavefunction:

$$\hat{\mathbf{p}} = -\mathrm{i}\hbar\nabla, \qquad \hat{E} = \mathrm{i}\hbar\frac{\partial}{\partial t} \tag{2.2}$$

to give

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r},t)\right]\psi(\mathbf{r},t) = \mathrm{i}\hbar\frac{\partial}{\partial t}\psi(\mathbf{r},t). \tag{2.3}$$

This equation is first order in the time derivatives and second order in the space derivatives which seems at odds with the spacetime concept of special relativity. We would expect a relativistic version of the Schrödinger equation to be of equal order in space and time derivatives.

The relativistic energy-momentum relationship is [27]

$$E^2 = c^2p^2 + (mc^2)^2 \tag{2.4}$$

and we could make the substitutions (2.2) to get the *Klein-Gordon equation*

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + m^2c^4 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} + V(\mathbf{r},t)\right]\psi(\mathbf{r},t) = 0 \tag{2.5}$$

which preserves the spacetime symmetry. The problem is that the probability density cannot be defined so that it is always positive [27]. Nevertheless, the Klein–Gordon equation does find a useful rôle in quantum field theory, describing mesons, which are bosons [28].

A second approach, taken by Dirac [29] is to construct a Hamiltonian which is first order in both space and time:

$$c\boldsymbol{\alpha}\cdot\mathbf{p} + \beta mc^2 = E, \tag{2.6}$$

where $\boldsymbol{\alpha}$ and β are determined by squaring (2.6) to recover the relativistic energy–momentum relationship (2.4). This leads to the conditions

$$\alpha_i\beta + \beta\alpha_i = 0, \quad \alpha_i^2 = 1, \quad \beta^2 = 1, \quad i = x, y, z, \tag{2.7}$$

which can be satisfied if $\boldsymbol{\alpha}$ and β are 4×4 matrices [30], suggesting that the wavefunction must now be a 4 component column vector. There are various representations of the $\boldsymbol{\alpha}$ and β matrices [27] but in the following, the *Dirac representation* (sometimes called the Dirac–Pauli representation) is used

$$\alpha_i = \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} \mathbb{I}_2 & 0 \\ 0 & -\mathbb{I}_2 \end{pmatrix}, \quad i = x, y, z, \tag{2.8}$$

where $\mathbb{I}_2$ is the 2×2 unit matrix and the σ_i are the Pauli spin matrices (1.12). The appearance of the Pauli matrices is encouraging as it hints that this Hamiltonian may include the electron spin. Replacing the momentum operator $\mathbf{p}$ with $-\mathrm{i}\hbar\nabla$ as before, we have the Dirac equation

$$\left[-\mathrm{i}c\hbar\boldsymbol{\alpha}\cdot\nabla + \beta mc^2 + V(\mathbf{r}) - ec\alpha\cdot\mathbf{A}(\mathbf{r}) - W\right]\psi(\mathbf{r}) = 0. \tag{2.9}$$

Here, $\mathbf{A}(\mathbf{r})$ is the vector potential due to an external magnetic field $\mathbf{B}(\mathbf{r}) = \nabla\times\mathbf{A}(\mathbf{r})$, e the electron charge, $V(\mathbf{r})$ is the external scalar potential and the energy is written

as W, rather than E, as it also contains the rest energy mc^2. The Dirac equation is consistent with special relativity as it is invariant under Lorentz transformations [10], [27], [5].

In the Schrödinger theory the probability density, $\rho(\mathbf{r},t)$ and current density, $\mathbf{j}(\mathbf{r},t)$ satisfy a continuity equation [9]

$$\frac{\partial}{\partial t}\rho(\mathbf{r},t)+\nabla\cdot\mathbf{j}(\mathbf{r},t)=0, \tag{2.10}$$

which is used to define the probability and current densities

$$\rho(\mathbf{r},t)=\int \mathrm{d}^3\mathbf{r}\,\psi^*(\mathbf{r},t)\psi(\mathbf{r},t), \qquad \mathbf{j}(\mathbf{r},t)=-\mathrm{i}\frac{\hbar}{2m}\int \mathrm{d}^3\mathbf{r}\left[\psi^*(\mathbf{r},t)\nabla\psi(\mathbf{r},t)-\psi(\mathbf{r},t)\nabla\psi^*(\mathbf{r},t)\right]. \tag{2.11}$$

An identical continuity equation can be derived for the Dirac equation [10] and used to define the probability density and current

$$\rho(\mathbf{r},t)=\int \mathrm{d}^3\mathbf{r}\,\psi^\dagger(\mathbf{r},t)\psi(\mathbf{r},t), \qquad \mathbf{j}(\mathbf{r},t)=c\int \mathrm{d}^3\mathbf{r}\,\psi^\dagger(\mathbf{r},t)\boldsymbol{\alpha}\psi(\mathbf{r},t). \tag{2.12}$$

Note that the complex conjugation of the non-relativistic theory has been replaced by Hermitian conjugation due to the 4-component nature of the Dirac wavefunction.

2.1.1 Plane Wave Solutions

As a first illustration of some of the interesting physics contained in the Dirac equation it seems sensible to look at the plane wave solutions for free particles. The relativistic energy momentum relationship is second order in the energy, $W^2 = c^2p^2+(mc^2)^2$. This means that we have two possible signs for the energy in (2.9), $W=\pm\sqrt{c^2p^2+(mc^2)^2}$. In the absence of external potentials, there are four linearly independent solutions to (2.9), two for each possible sign of the energy [30]. The upper two components of the solutions for the positive sign of the energy and the lower two for the negative sign are the two component spinors describing spin in the non-relativistic theory (1.11):

$$\phi_\uparrow=\begin{pmatrix}1\\0\end{pmatrix}, \qquad \phi_\downarrow=\begin{pmatrix}0\\1\end{pmatrix} \tag{2.13}$$

The other components can be written in terms of these spinors and the Pauli matrices so that free particle solutions may be expressed in the form

$$\psi^{(+)}_{m_s}(\mathbf{r})=\begin{bmatrix}\phi_{m_s}\\ \left(\dfrac{c\hbar}{W_++mc^2}\right)\boldsymbol{\sigma}\cdot\mathbf{k}\,\phi_{m_s}\end{bmatrix}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}, \quad \psi^{(-)}_{m_s}(\mathbf{r})=\begin{bmatrix}\left(\dfrac{c\hbar}{W_--mc^2}\right)\boldsymbol{\sigma}\cdot\mathbf{k}\,\phi_{m_s}\\ \phi_{m_s}\end{bmatrix}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}. \tag{2.14}$$

where $m_s = +1/2$ for spin up and $m_s = -1/2$ for spin down.

The four plane wave solutions can be separated by the sign of the energy and for each sign there are two solutions corresponding to the two different spins. The Dirac Hamiltonian was constructed to reflect the symmetries of spacetime, Lorentz invariance, and describes the spin of the electron. This demonstrates that spin is a relativistic phenomenon. It is worth remarking on the interpretation and consequences of the positive and negative energy solutions[‡]. Any negative energy state by definition has a lower energy than any positive energy state. If there is an empty negative state anywhere in the spectrum, then a positive energy electron should decay into it, emitting a photon. This raises the alarming consequence that the vacuum could suck all the electrons out of the universe and vanish in a burst of radiation. The way out of this is to assume that all the negative energy states are filled with electrons and all the positive states are empty. The two spectra for free particles are separated by a forbidden region of $2mc^2$ because a free particle cannot have an energy less than its rest energy. Below $-mc^2$ down to $-\infty$ all negative states are occupied and from mc^2 up to $+\infty$ all the usual positive electron states are empty. The Pauli exclusion principle explains why there can be no transitions to the filled negative states. If a negative energy electron is excited into a positive state, it leaves a hole in the negative spectrum. The hole will look like a particle with the same mass and equal and opposite charge, spin and momentum. Negative energy electrons can therefore be viewed as particles of the electron mass but with equal and opposite charge and spin (see figure 2-1). These are positrons. However, this definition of positrons leads to problems since the vacuum must be a state in which all allowed negative energies are occupied, leading to an infinite energy density. This difficulty is an indication that the single particle Dirac equation is insufficient for a correct description of relativistic quantum mechanics and was a catalyst to the development of the many–body picture of quantum field theory, which allows for the creation and destruction of particles. Nevertheless, many meaningful results are revealed without recourse to many–body quantum field theory and in this work we remain within the single particle picture where the total number of particles is conserved.

2.1.2 Non–Relativistic Limits

We have seen that if we let the speed of light, c, approach infinity we can investigate the non-relativistic limit and compare relativistic and non-relativistic quantities. To take the non-relativistic limit of the Dirac equation we write the four-component wavefunctions in as (2.9)

$$\psi(\mathbf{r}) = \begin{pmatrix} \phi(\mathbf{r}) \\ \chi(\mathbf{r}) \end{pmatrix} \tag{2.15}$$

[‡]We should make an important distinction here. Electrons in bound states of atoms are conventionally considered as having negative energies, but when we talk about negative energies here we are referring to particles which take the negative square root in the energy-momentum relationship

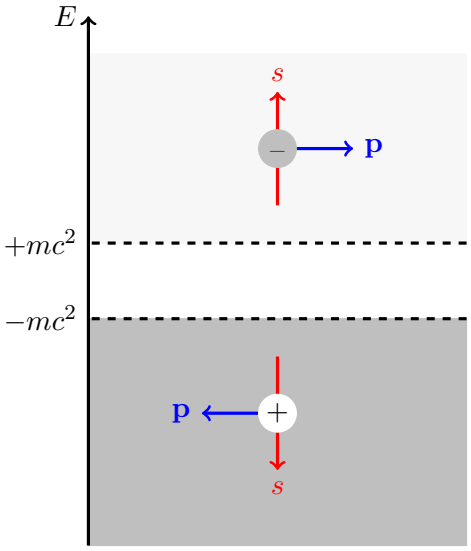

Figure 2-1: The Dirac picture of the vacuum. Negative energy states are filled up to $-mc^2$. The promotion of a negative energy particle to an empty positive state above $+mc^2$ creates a negatively charged electron of spin s and momentum $\mathbf{p}$. A hole is left behind of equal and opposite charge, spin and momentum; a positron.

and separate out the rest energy, setting $W = E+mc^2$, where E is the non-relativistic energy. The Dirac equation in this two component form is

$$\begin{pmatrix} V(\mathbf{r}) - E & c\boldsymbol{\sigma}\cdot(\hat{\mathbf{p}} - e\mathbf{A}(\mathbf{r})) \\ c\boldsymbol{\sigma}\cdot(\hat{\mathbf{p}} - e\mathbf{A}(\mathbf{r})) & V(\mathbf{r}) - E - 2mc^2 \end{pmatrix} \begin{pmatrix} \phi(\mathbf{r}) \\ \chi(\mathbf{r}) \end{pmatrix} = 0, \tag{2.16}$$

an equation for each of the two-component parts $\phi(\mathbf{r})$ and $\chi(\mathbf{r})$. Combining the two equations gives

$$c^2\boldsymbol{\sigma}\cdot(\hat{\mathbf{p}} - e\mathbf{A}(\mathbf{r}))\frac{\boldsymbol{\sigma}\cdot(\hat{\mathbf{p}} - e\mathbf{A}(\mathbf{r}))}{E - V(\mathbf{r}) + 2mc^2}\phi(\mathbf{r}) + (E - V(\mathbf{r}))\phi(\mathbf{r}) = 0. \tag{2.17}$$

At this point we introduce an identity for the Pauli matrices [10]

$$(\boldsymbol{\sigma}\cdot\mathbf{a})(\boldsymbol{\sigma}\cdot\mathbf{b}) = \mathbf{a}\cdot\mathbf{b} + \mathrm{i}\boldsymbol{\sigma}\cdot(\mathbf{a}\times\mathbf{b}) \tag{2.18}$$

where $\mathbf{a}$ and $\mathbf{b}$ are two vectors which commute. Letting $c \to \infty$ in (2.17) and writing the magnetic field $\mathbf{B}(\mathbf{r}) = \nabla\times\mathbf{A}(\mathbf{r})$ we obtain

$$\left[\frac{1}{2m}(\hat{\mathbf{p}} - e\mathbf{A}(\mathbf{r}))^2 - \frac{e\hbar}{2m}\boldsymbol{\sigma}\cdot\mathbf{B}(\mathbf{r}) + V(\mathbf{r}) - E\right]\phi(\mathbf{r}) = 0, \tag{2.19}$$

which is the Pauli equation (1.16) describing a non-relativistic spin-1/2 particle in a magnetic field (but we have now assumed no time dependence).

The factor of, $\mu = e\hbar/2m$, describing the electron magnetic moment has appeared naturally from the Dirac equation whereas it was added *ad hoc* in the non–

relativistic case. If $\mathbf{B}(\mathbf{r}) = 0$ in (2.19) then we have a separate Schrödinger equation for each component of $\phi(\mathbf{r})$ as there is no coupling of the spins, returning to the spin degeneracy of the non–relativistic theory.

It is also instructive to look at the non–relativistic limit of the free–particle solutions (2.14). As $c \to \infty$, $W_\pm \to \pm mc^2$ and we see that for electron-like, positive energy solutions the lower component is of order v/c times the upper component and the opposite is true for the negative-energy, positron like solutions. For this reason, the upper two components of an electron wavefunction are known as the *large component*, $\psi_l(\mathbf{r})$, and the lower two the *small component*, $\psi_s(\mathbf{r})$. The converse is true for positrons. In the absence of an external magnetic field there is a general relationship between the large and small components for electrons, from (2.16)

$$\psi_s(\mathbf{r}) = \frac{-\mathrm{i}\hbar c}{W - V(\mathbf{r}) + mc^2}\boldsymbol{\sigma} \cdot \nabla\psi_l(\mathbf{r}). \tag{2.20}$$

It is clear from (2.20) that the small component vanishes in the $c \to \infty$ limit and the large component reduces to the non–relativistic wavefunction. We will investigate the $c \to \infty$ limit frequently throughout the rest of this thesis in order to highlight both differences and similarities between the relativistic and non–relativistic quantities and expressions used in the relativistic embedding method.

2.1.3 Solutions for Spherically Symmetric Potentials

In addition to plane wave solutions for free particles, we will also require wavefunctions which are applicable to studying atoms and so consider the solutions of the Dirac equation with a spherically symmetric potential, $V(r)$. It is not quite as straightforward to formulate the Dirac equation in spherical polar coordinates as it is to do the same for the Schrödinger equation. This is because of the presence of the $\boldsymbol{\alpha}$ matrices in the kinetic energy term and the more complicated angular momentum considerations due to spin. In the absence of a magnetic vector potential $\mathbf{A}(\mathbf{r})$, the radial Dirac equation in spherical polar coordinates is [10]

$$\left[\mathrm{i}c\gamma_5\tilde{\sigma}_r\left(\hbar\frac{\partial}{\partial r} + \frac{\hbar}{r} - \frac{\beta\hat{K}}{r}\right) + \beta mc^2 + V(r) - W\right]\psi(\mathbf{r}) = 0, \tag{2.21}$$

where, $\mathbf{r} = (r, \theta, \phi)$, and we have introduced the following quantities

$$\tilde{\sigma}_r = \tilde{\boldsymbol{\sigma}} \cdot \hat{\mathbf{e}}_r, \qquad \tilde{\boldsymbol{\sigma}} = \begin{pmatrix} \boldsymbol{\sigma} & 0 \\ 0 & \boldsymbol{\sigma} \end{pmatrix}, \qquad \gamma_5 = \mathrm{i}\beta\alpha_x\beta\alpha_y\beta\alpha_z\beta. \tag{2.22}$$

The unit vector $\hat{\mathbf{e}}_r$ is in the radial direction. The operator $\hat{K}$ is related to the total angular momentum and is given by [31]

$$\hat{K} = \hat{\mathbf{L}} \cdot \tilde{\boldsymbol{\sigma}} + \hbar. \tag{2.23}$$

Eigenfunctions, $\Omega_\Lambda(\hat{\mathbf{r}})$ and eigenvalues, κ, of $\hat{K}$ satisfy

$$\hat{K}\Omega_\Lambda(\hat{\mathbf{r}}) = -\hbar\kappa\Omega_\Lambda(\hat{\mathbf{r}}), \tag{2.24}$$

where $\hat{\mathbf{r}} = (\theta, \phi)$. The eigenfunctions are orthogonal as one might expect:

$$\int_{4\pi} \mathrm{d}^2\hat{\mathbf{r}}\, \Omega^\dagger_\Lambda(\hat{\mathbf{r}})\Omega_{\Lambda'}(\hat{\mathbf{r}}) = \delta_{\Lambda\Lambda'}, \tag{2.25}$$

and are called the *spin-angular functions* [31]. The composite index $\Lambda \equiv (\kappa, \mu)$ where μ is equal to the non-relativistic azimuthal total angular momentum number m_j. There are definite relationships between κ and the non-relativistic l and j:

$$\begin{aligned} \kappa = -l - 1 = -j - \frac{1}{2} \quad &\text{for} \quad j = l + \frac{1}{2} \\ \kappa = l = j + \frac{1}{2} \quad &\text{for} \quad j = l - \frac{1}{2} \end{aligned} \quad . \tag{2.26}$$

It is also useful to define $\overline{\Lambda} = (-\kappa, \mu)$ and $\bar{l}$ which is the value of l associated with $-\kappa$:

$$\begin{aligned} \bar{l} = l + 1 = -\kappa \quad &\text{for} \quad \kappa < 0 \\ \bar{l} = l - 1 \quad &\text{for} \quad \kappa > 0 \end{aligned} \quad . \tag{2.27}$$

The spin angular functions are linear combinations of products of spherical harmonics, $Y_l^{\mu-m_s}(\hat{\mathbf{r}})$, and the two component spinors ϕ_{m_s} from (1.11):

$$\Omega_\Lambda(\hat{\mathbf{r}}) = \sum_{m_s} C_{\Lambda m_s} Y_l^{\mu-m_s}(\hat{\mathbf{r}})\phi_{m_s} \tag{2.28}$$

and the $C_{\Lambda m_s}$ are Clebsch-Gordan coefficients

$$C_{\Lambda+\frac{1}{2}} = -\mathrm{sgn}(\kappa)\sqrt{\frac{\kappa + \frac{1}{2} - \mu}{2\kappa + 1}}, \qquad C_{\Lambda-\frac{1}{2}} = \sqrt{\frac{\kappa + \frac{1}{2} + \mu}{2\kappa + 1}}. \tag{2.29}$$

A further useful relationship satisfied by the spin angular functions is

$$\tilde{\sigma}_r\Omega_\Lambda(\hat{\mathbf{r}}) = -\Omega_{\overline{\Lambda}}(\hat{\mathbf{r}}). \tag{2.30}$$

Having defined the radial Dirac equation and relativistic spin angular functions

we are in a position to propose a solution to (2.21) of the form

$$\psi_\Lambda(\mathbf{r}) = \begin{pmatrix} g_\kappa(r)\Omega_\Lambda(\hat{\mathbf{r}}) \\ \mathrm{i}f_\kappa(r)\Omega_{\overline{\Lambda}}(\hat{\mathbf{r}}) \end{pmatrix} \tag{2.31}$$

where $g_\kappa(r)$ and $f_\kappa(r)$ are functions satisfying the coupled first order differential equations [10], [32]

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}r}g_\kappa(r) &= -\frac{\kappa+1}{r}g_\kappa(r) + \frac{1}{c\hbar}(W - V(r) + mc^2)f_\kappa(r) \\ \frac{\mathrm{d}}{\mathrm{d}r}f_\kappa(r) &= \frac{\kappa-1}{r}f_\kappa(r) - \frac{1}{c\hbar}(W - V(r) - mc^2)g_\kappa(r) \end{aligned} \quad . \tag{2.32}$$

The factor of i in the lower component of (2.31) ensures that $g_\kappa(r)$ and $f_\kappa(r)$ are real for real W and $V(r)$.

Putting $V(r) = 0$ in (2.32) gives the free particle solutions

$$\psi^0_\Lambda(\mathbf{r}) = \begin{pmatrix} j_l(kr)\Omega_\Lambda(\hat{\mathbf{r}}) \\ \mathrm{i}\gamma k\,\mathrm{sgn}(\kappa)j_{\bar{l}}(kr)\Omega_{\overline{\Lambda}}(\hat{\mathbf{r}}) \end{pmatrix}, \tag{2.33}$$

where $\gamma = c\hbar/(W + mc^2)$ and $j_l(kr)$ is a spherical Bessel function chosen so that the solution is regular at the origin [33]. These can be shown to be equivalent to the plane waves (2.14) by using the following expansion of exponentials and two-component spinors [10]

$$\phi_{m_s}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} = 4\pi\sum_\Lambda \mathrm{i}^l C_{\Lambda m_s}\, j_l(kr)\left[Y_l^{\mu-m_s}(\hat{\mathbf{k}})\right]^*\Omega_\Lambda(\hat{\mathbf{r}}). \tag{2.34}$$

Following the free-particle solutions for the radial Dirac equation one would expect the solutions for the Coulomb potential to follow. Unfortunately, even compared to their non-relativistic counterparts, these solutions are extremely complicated. Fortunately they are not explicitly needed in this work, since the coupled equations can be solved numerically for a given potential and we will only say that they are written in terms of confluent hypergeometric functions [10]. However, since we have discussed the spin orbit interaction in non-relativistic quantum mechanics it will be useful to talk about the atomic solutions of the Dirac equation and the consequences of spin-orbit coupling.

2.1.4 The Spin-Orbit Interaction and the Dirac Equation

In non-relativistic quantum mechanics the spin-orbit interaction has to be included as an additional interaction. We have seen that it breaks the symmetry between orbital and spin angular momenta and destroys the degeneracy of spin up and spin down in atoms. Since it is a relativistic correction, we should be able to derive the

spin-orbit interaction term from the non-relativistic limit of the Dirac equation by retaining terms to order $1/c^2$.

Starting from (2.17) and neglecting any magnetic potentials we obtain

$$\frac{1}{2m}\boldsymbol{\sigma}\cdot\mathbf{p}\left(1+\frac{E-V(r)}{2mc^2}\right)^{-1}\boldsymbol{\sigma}\cdot\mathbf{p}\,\phi(\mathbf{r})+V(r)\phi(\mathbf{r})=E\phi(\mathbf{r}) \tag{2.35}$$

as an equation satisfied by the large component of the wavefunction.

Expanding the term in brackets to order $1/c^2$ yields

$$\frac{1}{2m}\boldsymbol{\sigma}\cdot\mathbf{p}\boldsymbol{\sigma}\cdot\mathbf{p}\phi(\mathbf{r})-\frac{1}{4m^2c^2}\boldsymbol{\sigma}\cdot\mathbf{p}\left[E-V(r)\right]\boldsymbol{\sigma}\cdot\mathbf{p}\phi(\mathbf{r})+V(r)\phi(\mathbf{r})=E\phi(\mathbf{r}). \tag{2.36}$$

From (2.18)

$$\begin{aligned}\boldsymbol{\sigma}\cdot\mathbf{p}\boldsymbol{\sigma}\cdot\mathbf{p}\phi(\mathbf{r}) &= -\hbar^2\boldsymbol{\sigma}\cdot\nabla\boldsymbol{\sigma}\cdot\nabla\phi(\mathbf{r})\\ &= -\hbar^2\nabla^2\phi(\mathbf{r})-\mathrm{i}\hbar^2\boldsymbol{\sigma}\cdot[\nabla\times\nabla\phi(\mathbf{r})]\\ &= -\hbar^2\nabla^2\phi(\mathbf{r})\end{aligned} \tag{2.37}$$

For the second term, noting that

$$\begin{aligned}&[\mathbf{p},V(r)]\,\phi(\mathbf{r})=(\mathbf{p}V(r))\,\phi(\mathbf{r})-V(r)\,(\mathbf{p}\phi(\mathbf{r}))\\ \Rightarrow\qquad &\mathbf{p}V(r)=V(r)\mathbf{p}-\mathrm{i}\hbar\nabla V(r),\end{aligned} \tag{2.38}$$

and using the properties of the $\boldsymbol{\sigma}$ matrices again

$$\boldsymbol{\sigma}\cdot(\nabla V(r))\,\boldsymbol{\sigma}\cdot(\mathbf{p}\phi(\mathbf{r}))=\nabla V(r)\cdot\nabla\phi(\mathbf{r})+\mathrm{i}\boldsymbol{\sigma}\cdot[\nabla V(r)\times\nabla\phi(\mathbf{r})]\,. \tag{2.39}$$

Substituting these back into (2.36) we have

$$\begin{aligned}&-\frac{\hbar^2}{2m}\left[1-\frac{E-V(r)}{2mc^2}\right]\nabla^2\phi(\mathbf{r})-\frac{\hbar^2}{4m^2c^2}\nabla V(r)\cdot\nabla\phi(\mathbf{r})\\ &\qquad+\frac{\hbar^2}{4m^2c^2}\boldsymbol{\sigma}\cdot[\nabla V(r)\times\mathbf{p}]+V(r)\phi(\mathbf{r})=E\phi(\mathbf{r})\end{aligned} \tag{2.40}$$

Since the potential only depends on the radial coordinate we can put [34]

$$\nabla V(r)\cdot\nabla\phi(\mathbf{r})=\frac{\mathrm{d}V(r)}{\mathrm{d}r}\frac{\partial\phi(\mathbf{r})}{\partial r},\qquad \nabla V(r)=\frac{1}{r}\frac{\mathrm{d}V(r)}{\mathrm{d}r}\mathbf{r}. \tag{2.41}$$

The final step is to notice that, $(E-V(r))$ is equal to the kinetic energy, which in the non-relativistic limit we can write as $\mathbf{p}^2/2m$. The non-relativistic limit of the Dirac

equation for a spherical potential, retaining corrections to order $1/c^2$ is therefore

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(r) - E\right]\phi(\mathbf{r}) - \frac{\mathbf{p}^4}{8m^3c2}\phi(\mathbf{r})$$

$$-\frac{\hbar^2}{4m^2c^2}\frac{\mathrm{d}V(r)}{\mathrm{d}r}\frac{\mathrm{d}\phi(\mathbf{r})}{\mathrm{d}r} + \frac{\hbar^2}{4m^2c^2}\frac{1}{r}\frac{\mathrm{d}V(r)}{\mathrm{d}r}\boldsymbol{\sigma}\cdot[\mathbf{r}\times\mathbf{p}]\,\phi(\mathbf{r}) = 0. \tag{2.42}$$

The first term is the non-relativistic Schrödinger equation and the second is the first order kinetic energy correction. The third term is the Darwin term and has no non-relativistic analog [10]. The final term

$$\frac{\hbar^2}{4m^2c^2}\frac{1}{r}\frac{\mathrm{d}V(r)}{\mathrm{d}r}\boldsymbol{\sigma}\cdot[\mathbf{r}\times\mathbf{p}] \tag{2.43}$$

is more recognisable after making some substitutions. The $\mathbf{r}\times\mathbf{p}$ is the orbital angular momentum $\mathbf{L}$ and the spin operator is $\mathbf{S} = \hbar\boldsymbol{\sigma}/2$. The final term is therefore the spin orbit interaction to first order in $1/c^2$

$$\frac{1}{2m^2c^2}\frac{1}{r}\frac{\mathrm{d}V(r)}{\mathrm{d}r}\mathbf{L}\cdot\mathbf{S}. \tag{2.44}$$

Having established that the Dirac equation contains the spin-orbit coupling term, we can use it to clarify the relativistic angular momentum quantum numbers $\Lambda = (\kappa, \mu)$. The orbital, $\mathbf{L}^2, L_z$ and spin, $\mathbf{S}^2, S_z$, angular momentum operators commute with the Schrödinger Hamiltonian in the absence of spin-orbit coupling meaning that states with $(n, l, m_l, m_s = +1/2)$ and $(n, l, m_l, m_s = -1/2)$ are degenerate. If the spin-orbit interaction is included, $\mathbf{L}$ and $\mathbf{S}$ do not commute separately but the total angular momentum $\mathbf{J}^2, J_z$ does and the degeneracy is lifted. It can be shown [10] that $\mathbf{L}^2, L_z$ and $\mathbf{S}^2, S_z$ do not commute with the Dirac Hamiltonian but $\mathbf{J}^2, J_z$ does. The operator, $K = \mathbf{L}\cdot\boldsymbol{\sigma} + \hbar$ arises naturally in the Dirac equation for spherically symmetric potentials, rather than $\mathbf{J}$. This is the reason for the use of K and its associated quantum numbers, $\Lambda = (\kappa, \mu)$, over the $\mathbf{J}$ encountered in non-relativistic atoms. Again, it can be shown that [10]

$$\left[K, \mathbf{J}^2\right] = [K, J_z] = 0, \qquad \left[\mathbf{J}^2, H_D\right] = \left[K^2, H_D\right], \qquad [K, H_D] = 0, \tag{2.45}$$

where H_D is the Dirac Hamiltonian.

The energy of atomic energy levels of Hydrogenic atoms, to first order in the relativistic case, is [8]

$$E = -\frac{Z^2me^4}{2(4\pi\varepsilon_0\hbar)^2}\frac{1}{n^2} + \frac{Z^4me^8}{2(4\pi\varepsilon_0\hbar)^4c^2}\frac{1}{n^4}\left(\frac{n}{j+\frac{1}{2}} - \frac{3}{4}\right) \tag{2.46}$$

and the non-relativistic expression, including the spin orbit correction, is reproduced

for comparison

$$E_{NR} = -\frac{Z^2 me^4}{2(4\pi\varepsilon_0\hbar)^2}\frac{1}{n^2} + \frac{Z^4 me^8}{2(4\pi\varepsilon_0\hbar)^4 c^2}\frac{j(j+1) - l(l+1) - \frac{3}{4}}{n^3 l(l+1)(2l+1)}. \tag{2.47}$$

The first term in both expressions is clearly equivalent and gives the energy levels of the non-relativistic atom. The second term in the non-relativistic energy is purely the first order correction due to the spin orbit interaction from perturbation theory. A direct comparison with the final term in the relativistic expression is difficult because this contains not just the spin-orbit correction but the kinetic energy and Darwin corrections. However, it demonstrates the dependence only on the total angular momentum, j, that we would expect from the symmetries and commutation relations discussed above.

The main point of this section is to illustrate that spin is a natural consequence of special relativity and is implicit in the Dirac equation, in contrast to its *ad hoc* inclusion in non-relativistic theory. This means that we do not need to treat the spin-orbit interaction as a perturbation; it is always included in solutions of the Dirac equation to all orders, not just to order $1/c^2$. In addition, solutions of the Dirac equation also include all of the other relativistic effects to all orders, including the mass-velocity corrections and others which have no non-relativistic counterparts.

2.1.5 The Koelling-Harmon Equation

The Schrödinger equation contains no relativistic effects and the Dirac equation contains all relativistic effects to all orders, including the mass-velocity corrections and the spin orbit interaction. The main problem with including the spin-orbit interaction is that breaks the symmetry between spin and orbital angular momentum and l, m_l and m_s are no longer good quantum numbers; we must label the states by the total angular momentum j, m_j. Since the mass-velocity corrections produce significant effects such as the lowering of s state energies and raising of d state energies, it would be useful to have an 'intermediate' relativistic equation which includes the mass-velocity effects and the Darwin correction, but not the spin orbit term. This is the Koelling-Harmon equation, sometimes called the *scalar relativistic approximation* [35].

The solution of the Dirac equation for a spherically symmetric potential is (2.31)

$$\psi_\Lambda(\mathbf{r}) = \begin{pmatrix} g_\kappa(r)\Omega_\Lambda(\hat{\mathbf{r}}) \\ \mathrm{i} f_\kappa(r)\Omega_{\overline{\Lambda}}(\hat{\mathbf{r}}) \end{pmatrix} \tag{2.48}$$

where $g_\kappa(r)$ and $f_\kappa(r)$ are solutions of the coupled equations (2.32)

$$\frac{\mathrm{d}}{\mathrm{d}r}f_\kappa(r) = \frac{1}{c}(V-E)g_\kappa(r) + \left(\frac{\kappa-1}{r}\right)f_\kappa(r) \tag{2.49}$$

$$\frac{\mathrm{d}}{\mathrm{d}r}g_\kappa(r) = -\left(\frac{\kappa+1}{r}\right)g_\kappa(r) + 2Mcf_\kappa(r) \tag{2.50}$$

where $E = W - mc^2$ and

$$M = m + \frac{1}{2c^2}(E-V). \tag{2.51}$$

Combining the equations for $g_\kappa(r)$ and $f_\kappa(r)$ gives a second order equation for the large component

$$-\frac{1}{2M}\left[\frac{\mathrm{d}^2}{\mathrm{d}r^2}g_\kappa(r) + \frac{2}{r}\frac{\mathrm{d}}{\mathrm{d}r}g_\kappa(r) - \frac{l(l+1)}{r^2}g_\kappa(r)\right] - \frac{1}{4M^2c^2}\frac{\mathrm{d}}{\mathrm{d}r}V(r)\frac{\mathrm{d}}{\mathrm{d}r}g_\kappa(r)$$

$$+V(r)g_\kappa(r) - \frac{1}{4M^2c^2}\left(\frac{\kappa-1}{r}\right)\frac{\mathrm{d}}{\mathrm{d}r}V(r)g_\kappa(r) = Eg_\kappa(r). \tag{2.52}$$

The last term on the left-hand side is the spin orbit interaction and is the only term which depends on the sign of κ, whether $j = l \pm 1/2$, coupling the spin and orbital angular momenta.

By defining the function

$$\chi_\kappa(r) = \frac{1}{2Mc}\frac{\mathrm{d}}{\mathrm{d}r}g_\kappa(r) = f_\kappa(r) - \frac{1}{2Mc}\left(\frac{1+\kappa}{r}\right)g_\kappa(r) \tag{2.53}$$

and dropping the spin orbit term, (2.52) becomes

$$\frac{\mathrm{d}}{\mathrm{d}r}\chi_l(r) = -\frac{2}{r}\chi_l(r) + \left[\frac{l(l+1)}{2Mcr^2} + \frac{1}{c}(E-V(r))\right]g_l(r) \tag{2.54}$$

where the index κ has been replaced by l since there is no longer any dependence on j because the spin-orbit coupling term has been removed. The wavefunction now depends on only the non relativistic quantum numbers l, m_l, m_s.

The Koelling Harmon approach is usually implemented by defining the functions

$$p_l(r) = rg_l(r), \qquad q_l(r) = rc\chi_l, \tag{2.55}$$

giving the coupled equations for $p_l(r)$ and $q_l(r)$

$$\frac{\mathrm{d}}{\mathrm{d}r}p_l(r) = 2Mq_l(r) + \frac{1}{r}p_l(r) \tag{2.56}$$

$$\frac{\mathrm{d}}{\mathrm{d}r}q_l(r) = -\frac{1}{r}q_l(r) + \left[\frac{l(l+1)}{2Mr^2} + (V(r) - E)\right]p_l(r). \tag{2.57}$$

The scalar relativistic approximation is widely used in electronic structure calculations as it allows the inclusion of relativistic mass-velocity corrections without extra computational cost.

Before moving on it is necessary at this point to clarify what we mean from now on when we talk about 'relativistic' and 'non-relativistic'. When we say *relativistic* we *always* mean the full Dirac equation. Non-relativistic means either the Schrödinger equation or the Koelling Harmon/scalar relativistic equation. When a distinction is needed between these two they will be named specifically.

In the final part of this section we will discuss a further class of solutions particular to the Dirac equation which will be useful later on.

2.1.6 Left and Right Solutions and Time Reversal

The different symmetries underlying a relativistic description of spacetime have already been mentioned when discussing the Lorentz transformations and the spin of the electron. We saw that the negative energy states could be described by a positron, the electron's anti-particle. This symmetry between matter and antimatter is fundamental to relativistic quantum mechanics. An important symmetry of the Dirac equation is CPT symmetry; Charge conjugation, Parity and Time-reversal [27], [10]. Charge conjugation is changing the sign of the charge on a particle, parity describes the symmetries of space inversion, $\mathbf{r} \to -\mathbf{r}$ and time reversal the symmetries as $t \to -t$. If we charge conjugate the Dirac equation for an electron we end up with the Dirac equation for an electron with opposite charge, spin and momentum. This is consistent with the positron description of negative energy states. We will talk about time reversal in more detail in a moment but here we will say that it is a symmetry of the Dirac equation which reverses the magnetic field and spin of the electron, consistent with the Kramers degeneracy discussed in connection with Rashba splitting at surfaces. The parity operation introduces a phase factor to the electron or positron wavefunction, leaving observable quantities unchanged. Applying all of these symmetry operations consecutively (CPT) has an interesting result which demonstrates the symmetry between matter and antimatter, whereby we effectively end up with a description of a positron being a negative energy electron moving backwards through spacetime! A result of this is that the energy spectrum of an antimatter atom (a positron orbiting an anti-proton for anti-hydrogen) is iden-

tical to that of a normal atom. Charge conjugation, parity and CPT are only really relevant in this work in the way that they reinforce the interpretation of the positive energy solutions of the Dirac equation as electrons. However, time-reversal symmetry should be investigated further since it appears in expressions for the Green's function and in investigations of ballistic transport.

The Dirac Hamiltonian is a 4×4 matrix, the wavefunction is a 4×1 spinor and the differential operator acts to a wavefunction on the right hand side:

$$\left[-\mathrm{i}c\hbar\boldsymbol{\alpha} \cdot \overrightarrow{\nabla} + \beta mc^2 + V(\mathbf{r}) - ec\boldsymbol{\alpha} \cdot \mathbf{A}(\mathbf{r}) - W\right] \psi(\mathbf{r}) = 0. \tag{2.58}$$

It is possible, however to define a Dirac equation in which the 4×4 operators act to the left on a wavefunction which is a 1×4 spinor [36]:

$$\psi^{\times}(\mathbf{r}) \left[-\mathrm{i}c\hbar\boldsymbol{\alpha} \cdot \overleftarrow{\nabla} + \beta mc^2 + V(\mathbf{r}) - ec\boldsymbol{\alpha} \cdot \mathbf{A}(\mathbf{r}) - W\right] = 0. \tag{2.59}$$

The kinetic energy operator $-\mathrm{i}c\hbar\boldsymbol{\alpha} \cdot \nabla$ must be Hermitian because the kinetic energy is an observable, so for some general function $f(\mathbf{r})$, [37]

$$-\mathrm{i}f(\mathbf{r})\overleftarrow{\nabla} = \mathrm{i}\overrightarrow{\nabla}f(\mathbf{r}), \tag{2.60}$$

where the arrows indicate the direction in which the gradient operator acts. Using (2.60) in (2.59) to make the gradient operator act to the right and Hermitian conjugating we have

$$\left[-\mathrm{i}c\hbar\boldsymbol{\alpha} \cdot \overrightarrow{\nabla} + \beta mc^2 + V^*(\mathbf{r}) - ec\boldsymbol{\alpha} \cdot \mathbf{A}^*(\mathbf{r}) - W^*\right] \psi^{\times\dagger}(\mathbf{r}) = 0 \tag{2.61}$$

where we have assumed the general case that both the scalar and vector potential and the total energy are complex. Left and right wavefunctions are related by some operation which transforms (2.61) into (2.59). To see this we introduce the relativistic time-reversal operator [38],

$$\hat{\tau} = -\mathrm{i}\tilde{\sigma}_y\tau_0, \tag{2.62}$$

where τ_0 is an operator which complex conjugates everything to its right and

$$\tilde{\sigma}_y = \begin{pmatrix} \sigma_y & 0 \\ 0 & \sigma_y \end{pmatrix}. \tag{2.63}$$

The operator $\hat{\tau}$ has the following properties [9],

$$\hat{\tau}^2 = -1, \qquad \hat{\tau}\hat{\tau}^\dagger = 1, \qquad \hat{\tau}\boldsymbol{\alpha}\hat{\tau}^\dagger = -\boldsymbol{\alpha}, \qquad \hat{\tau}\beta\hat{\tau}^\dagger = \beta, \qquad \hat{\tau}Z\tau = Z^* \tag{2.64}$$

where Z is complex. Applying this to (2.61),

$$\left[-\hat{\tau}\mathrm{i}\hat{\tau}^{\dagger}\, c\hbar\,\hat{\tau}\boldsymbol{\alpha}\hat{\tau}^{\dagger}\cdot\nabla+\hat{\tau}\beta\hat{\tau}^{\dagger}\, mc^2+\hat{\tau}V^{*}(\mathbf{r})\hat{\tau}^{\dagger}-ec\,\hat{\tau}\boldsymbol{\alpha}\hat{\tau}^{\dagger}\cdot\hat{\tau}\mathbf{A}^{*}(\mathbf{r})\hat{\tau}^{\dagger}-\hat{\tau}W^{*}\hat{\tau}^{\dagger}\right]\hat{\tau}\psi^{\times\dagger}(\mathbf{r})=0 \tag{2.65}$$

and using the properties of $\hat{\tau}$ we have

$$\left[-\mathrm{i}c\hbar\boldsymbol{\alpha}\cdot\overrightarrow{\nabla}+\beta mc^2+V(\mathbf{r})+ec\boldsymbol{\alpha}\cdot\mathbf{A}(\mathbf{r})-W\right]\psi^{\times\dagger}(\mathbf{r})=0 \tag{2.66}$$

which is identical to (2.58) if $\psi(\mathbf{r})=\hat{\tau}\psi^{\times\dagger}(\mathbf{r})$,except that the vector potential has the opposite sign. If we operate with $\hat{\tau}$ on a plane wave solution (2.14) we see that the spin is flipped.

The vector potential changes sign under the action of $\hat{\tau}$ because magnetic fields are caused by moving charges, which move in the opposite direction under time reversal ($\mathbf{B}$ is an axial vector). Angular momentum is also an axial vector and, since the electron spin is an angular momentum, a time reversed electron has its spin flipped.

Left, $\psi^{\times}$, and right, ψ, wavefunctions are therefore related by time-reversal and Hermitian conjugation

$$\psi^{\times}_{m_s}(\mathbf{r};\mathbf{A})=\left[\hat{\tau}\psi_{-m_s}(\mathbf{r};-\mathbf{A})\right]^{\dagger}, \tag{2.67}$$

If $\mathbf{A}(\mathbf{r})=0$ and $V(\mathbf{r})$ and W are real then (2.67) reduces to $\psi^{\times}(\mathbf{r})=\psi^{\dagger}(\mathbf{r})$.

For the Schrödinger equation in the absence of spin, time reversal is achieved through application of the complex conjugation operator τ_0 alone [9]. When comparing non–relativistic expressions which involve time reversed states we would therefore expect to encounter left hand solutions of the Dirac equation performing the role of complex conjugated wavefunctions of the Schrödinger equation. We will see this correspondence when discussing the spectral representation of the Green's function in the next section and in the context of electron transport in following chapters.

The time reversal operator $\hat{\tau}$ can be used to demonstrate the Kramers degeneracy mentioned in connection with the Rashba effect. Consider a free particle, represented by a plane wave, moving in the z-direction with momentum k_z,

$$\psi(k_z,m_s)=\begin{pmatrix}\phi_{m_s}\\ \gamma\sigma_z k_z\phi_{m_s}\end{pmatrix}\mathrm{e}^{\mathrm{i}k_z z} \tag{2.68}$$

The complex conjugation part of $\hat{\tau}$ only acts on the exponential since we assume that the energy and momentum are real. The effect of the time reversal operator on the spinors ϕ_{m_s} is easily shown to be

$$\hat{\tau}\phi_{m_s}=\mathrm{sgn}(m_s)\phi_{-m_s}, \tag{2.69}$$

and for the small component, using the anticommuting properties of the Pauli matrices, $\sigma_y\sigma_z = -\sigma_z\sigma_y$, one gets

$$\hat{\tau}\psi(k_z, m_s) = \mathrm{sgn}(m_s) \begin{pmatrix} \phi_{-m_s} \\ -\gamma\sigma_z k_z \phi_{-m_s} \end{pmatrix} \mathrm{e}^{-\mathrm{i}k_z z} = \mathrm{sgn}(m_s)\psi(-k_z, -m_s) \tag{2.70}$$

showing that the time reversal operator reverses the momentum and spin of the particle. The spin dependent phase factor, $\mathrm{sgn}(m_s)$ vanishes when taking matrix elements so that the time reversed particles are degenerate.

2.2 Green's Functions

Green's functions are found in almost all areas of Mathematical Physics and are a particularly powerful tool in condensed matter and electronic structure calculations [39], [40], [41], [27], [42], [43], [2], [44]. Physically, the Green's function at energy E, $G(\mathbf{r}, \mathbf{r}'; E)$, is a type of propagator, describing the response of a system at $\mathbf{r}$ to an impulse at $\mathbf{r}'$. It is therefore a natural quantity for describing scattering in solids [10] and is fundamental to the KKR method for electronic structure calculations [45]. More importantly for this work, they are central in the embedding method. They also provide a means of solving inhomogeneous differential equations and allow certain physical quantities to be calculated in a more convenient way. In this section we will present a general definition of the Green's function, apply this to the Schrödinger and Dirac theories and discuss some of its properties and representations which will be useful in the embedding method.

To introduce the general Green's function we consider a linear differential operator, $\mathcal{L}$ which transforms a vector $\mathbf{x}$ into some other vector, $\mathbf{y}$ in a vector space, L

$$\mathcal{L}\mathbf{x} = \mathbf{y}, \qquad \mathbf{x}, \mathbf{y} \in L \tag{2.71}$$

The inverse operator, $\mathcal{L}^{-1}$ transforms $\mathbf{y}$ back into $\mathbf{x}$

$$\mathcal{L}^{-1}\mathbf{y} = \mathbf{x}, \qquad \mathbf{x}, \mathbf{y} \in L. \tag{2.72}$$

This inverse operator is the Green's function [41] and

$$\mathcal{L}\mathcal{L}^{-1} = \mathbb{I}_n, \tag{2.73}$$

where $\mathbb{I}_n$ is the $n \times n$ unit matrix, n being the dimension of L. Both the Schrödinger and Dirac Hamiltonians are linear differential operators on the space of wavefunctions and we can write the Green's function at energy E, $G(E)$, as

$$G(E) = \frac{1}{E - H}. \tag{2.74}$$

In the position representation we can write (2.73) as

$$[H - E]\, G(\mathbf{r}, \mathbf{r}'; E) = -\delta(\mathbf{r} - \mathbf{r}') \tag{2.75}$$

where the $-$sign comes from the convention of writing $H - E$ rather than $E - H$ [46], [3].

As an example of the use of Green's functions we consider a system described by the Hamiltonian H_0, with wavefunction ϕ at energy E

$$[H_0 - E]\, \phi(\mathbf{r}) = 0. \tag{2.76}$$

If we now introduce an additional potential $V(\mathbf{r})$ we can find the new wavefunction ψ by solving

$$[H_0 - E]\, \psi(\mathbf{r}) = V(\mathbf{r})\psi(\mathbf{r}). \tag{2.77}$$

This situation frequently arises in scattering problems [10], [30]. The Green's function defined by (2.75) allows us to write

$$\psi(\mathbf{r}) = \phi(\mathbf{r}) + \int \mathrm{d}^3\mathbf{r}'\, G_0(\mathbf{r}, \mathbf{r}'; E) V(\mathbf{r}')\psi(\mathbf{r}'). \tag{2.78}$$

This is the Dyson equation for the wavefunction $\psi(\mathbf{r})$ [47].

2.2.1 The Spectral Representation of the Green's Function

The definition of the Green's function (2.75) suggests that it can be written in terms of the eigenfunctions of H. In the non-relativistic case we do this by writing the Green's function as

$$G(\mathbf{r}, \mathbf{r}'; E) = \sum_{nn'} a_{nn'} \phi_n(\mathbf{r}) \phi^*_{n'}(\mathbf{r}') \tag{2.79}$$

where $\phi_n(\mathbf{r})$ are the eigenfunctions of H and $A_{nn'}$ are some expansion coefficients. Substituting this into (2.75) and using the orthonormality of the eigenfunctions

$$\sum_{nn'} \phi_n(\mathbf{r}) \phi^*_{n'}(\mathbf{r}') = \delta_{nn'} \delta(\mathbf{r} - \mathbf{r}') \tag{2.80}$$

gives the spectral representation of the Green's function

$$G(\mathbf{r}, \mathbf{r}'; E) = \sum_{n} \frac{\phi_n(\mathbf{r}) \phi^*_n(\mathbf{r}')}{E - E_n}. \tag{2.81}$$

To illustrate the significance of the spectral representation we will consider a simple, one dimensional model system. Suppose we have a particle travelling from $-\infty$ towards some barrier $V(z)$ where $V(z) \to 0$ outside the limits Z_1 and Z_2. We will expect the wavefunction on the left hand side to consist of an incident and reflected

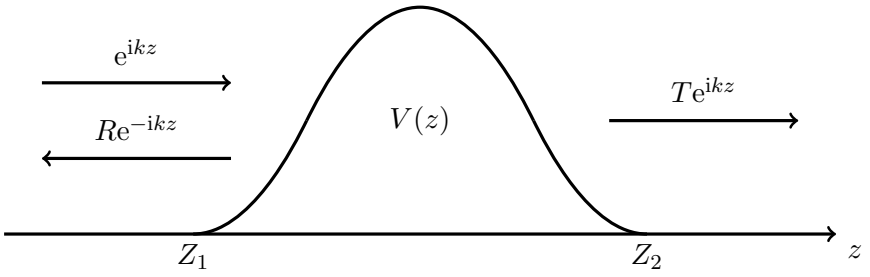

Figure 2-2: General 1-dimensional scattering from a potential $V(z)$ which is non-zero between Z_1 and Z_2. The boundary conditions on the wavefunction are that it should consist of an incident and reflected wave for $z < Z_1$ and a transmitted wave for $z > Z_2$.

wave and on the right hand side to be a transmitted wave (see figure 2-2). We want to solve the Dyson equation for the problem

$$\psi(z) = \phi(z) + \int_{Z_1}^{Z_2} \mathrm{d}z'\, G_0(z, z'; E) V(z') \psi(z') \tag{2.82}$$

where $\phi(z)$ is the free particle solution

$$\phi(z) = \mathrm{e}^{\mathrm{i}kz}. \tag{2.83}$$

and $G_0(z, z'; E)$ is the free particle Green's function, given in the spectral representation by

$$G_0(z, z'; k_0) = \frac{2m}{\hbar^2} \int_{-\infty}^{+\infty} \frac{\mathrm{d}k}{2\pi} \frac{\mathrm{e}^{\mathrm{i}k(z-z')}}{k_0^2 - k^2}. \tag{2.84}$$

We have an integral because the eigenvalue spectrum, k, is continuous and we have replaced the energy E with $\hbar^2 k_0^2/2m$. This highlights a problem; there are poles at $k = \pm k_0$. To do the integral we can add an infinitesimally small imaginary part to the energy of the Green's function, $k_0 \to k_0 + \mathrm{i}\eta$ and displace the poles from the real axis

$$G_0(z, z'; k_0 + \mathrm{i}\eta) = G_0^+(z, z'; k_0) = \frac{2m}{\hbar^2} \lim_{\eta \to 0} \int_{-\infty}^{+\infty} \frac{\mathrm{d}k}{2\pi} \frac{\mathrm{e}^{\mathrm{i}k(z-z')}}{(k_0 + \mathrm{i}\eta - k)(k_0 + \mathrm{i}\eta + k)}, \tag{2.85}$$

where the superscript + indicates that we have added an infinitesimal imaginary part to the energy. The integral can now be performed around a contour in the complex plane. If $z > z'$ we are on the right hand side of the scattering barrier and we need the wavefunction to be well behaved as $|k| \to +\infty$ to describe the transmitted wave travelling to the right. The contour is therefore in the upper half plane and contains the pole at $k = k_0 + \mathrm{i}\eta$ (see figure 2-3). For $z < z'$ we want to describe the reflected wave travelling towards and the contour is in the lower half plane enclosing the pole

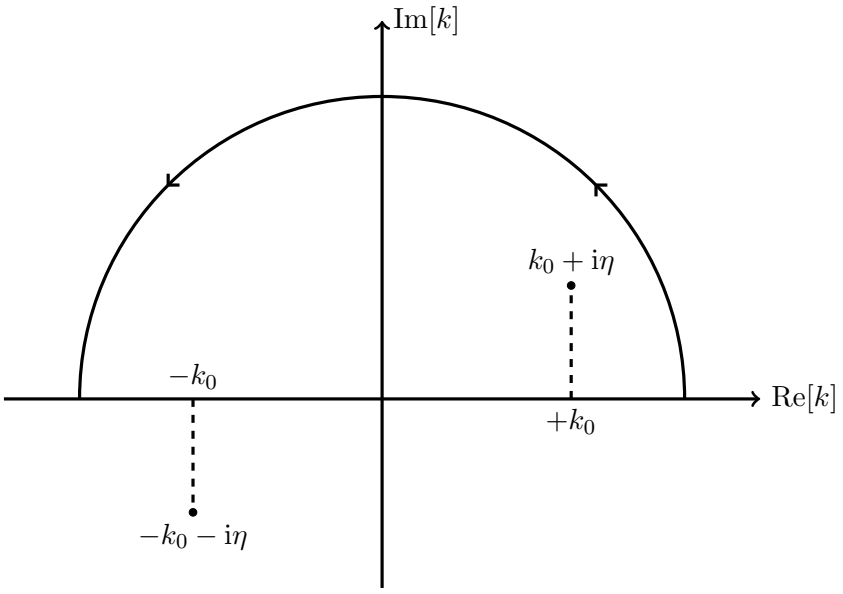

Figure 2-3: The anticlockwise contour in the complex k plane for performing the integral in (2.85), enclosing the pole at $k = k_0 + i\eta$ and ensuring the outgoing boundary condition on the wavefunction.)

at $k = -k_0 - i\eta$. Combining these integrals gives [34]

$$G_0^+(z, z'; k_0) = -\frac{i\hbar^2}{2mk_0} e^{ik_0|z-z'|}. \tag{2.86}$$

Substituting this back into the Dyson equation we have

$$\psi(z) = \phi(z) - \frac{i\hbar^2}{2mk_0} \int_{Z_1}^{Z_2} dz' e^{ik_0|z-z'|} \psi(z') V(z') \tag{2.87}$$

If $z > Z_2$ and we are to the right of the barrier, $\phi(z) = e^{ik_0 z}$ and

$$\psi(z) = e^{ik_0 z} \left[1 - \frac{i\hbar^2}{2mk_0} \int_{Z_1}^{Z_2} dz' e^{-ik_0 z'} \psi(z') V(z') \right] \tag{2.88}$$

which is a wave travelling towards $+\infty$ and the term in square brackets may be identified as a transmission coefficient. For $z < Z_1$

$$\psi(z) = e^{ik_0 z} + \frac{i\hbar^2}{2mk_0} e^{-ik_0 z} \int_{Z_1}^{Z_2} dz' e^{ik_0 z'} \psi(z') V(z') \tag{2.89}$$

which describes an incident wave travelling towards the barrier along with a reflected wave, the integral now representing the reflection coefficient. This simple example illustrates that we can include the correct boundary conditions for the wavefunction in the Green's function by adding a small positive imaginary part to the energy. This corresponds to an *outgoing* boundary condition.

Adding a small imaginary part to the energy in the spectral representation also allows us to derive a useful result. Using the identity [34]

$$\lim_{\eta\to 0}\frac{1}{x-a+\mathrm{i}\eta}=\frac{1}{x-a}-\mathrm{i}\pi\delta(x-a),\qquad x,a,\eta\in\mathbb{R} \tag{2.90}$$

the spectral representation gives [10]

$$-\lim_{\eta\to 0}\frac{1}{\pi}\mathrm{Im}\,G(\mathbf{r},\mathbf{r};E+\mathrm{i}\eta)=\sum_n|\psi_n(\mathbf{r})|^2\delta(E-E_n). \tag{2.91}$$

The right hand side is the local density of states, $n(\mathbf{r};E)$ therefore

$$n(\mathbf{r};E)=-\frac{1}{\pi}\mathrm{Im}G^+(\mathbf{r},\mathbf{r};E). \tag{2.92}$$

The density of states in a given volume can be found by calculating the local density of states and integrating over the volume. It is especially useful for systems such as surfaces where the density of states near the surface differs from that of the bulk crystal.

The Green's function can also be used to obtain the charge density $\rho(\mathbf{r})$ by a similar argument [10]:

$$\rho(\mathbf{r})=-\frac{e}{\pi}\mathrm{Im}\int_{-\infty}^{E_f}\mathrm{d}E\,G^+(\mathbf{r},\mathbf{r},E+\mathrm{i}\eta) \tag{2.93}$$

where E_f is the Fermi energy.

The spectral representation and expressions for the charge density and density of states may be extended to the relativistic case. The Dirac Green's function satisfies

$$[H_D-W]\,G(\mathbf{r},\mathbf{r}';W)=-\delta(\mathbf{r}-\mathbf{r}') \tag{2.94}$$

where H_D is the Dirac Hamiltonian in (2.9) and the spectral representation is [36]

$$G^+(\mathbf{r},\mathbf{r}';W+\mathrm{i}\eta)=\sum_n\frac{\phi_n(\mathbf{r})\otimes\phi_n^\times(\mathbf{r}')}{W-W_n} \tag{2.95}$$

where $\phi^\times(\mathbf{r})$ is a left-hand solution of the Dirac equation. The outer product of two vectors [48], $\otimes$ has been introduced such that, for a $N\times 1$ vector $\mathbf{a}$ and a $1\times N$ vector $\mathbf{b}$

$$\mathbf{a}\otimes\mathbf{b}=\begin{pmatrix}a_1\\a_2\\\vdots\\a_N\end{pmatrix}\otimes\begin{pmatrix}b_1&b_2&\cdots&b_N\end{pmatrix}=\begin{pmatrix}a_1b_1&a_1b_2&\cdots&a_1b_N\\a_2b_1&a_2b_2&\cdots&a_2b_N\\\vdots&&\ddots&\\a_Nb_1&a_Nb_2&&a_Nb_N\end{pmatrix}. \tag{2.96}$$

From now on we will not include the symbol $\otimes$ for the outer product since it should be clear when it is implied. From the spectral representation it is easy to see that the Dirac Green's function is a 4×4 matrix.

The corresponding relativistic expressions for the charge density, $\rho(\mathbf{r})$ and density of states, $n(W)$ are

$$\begin{aligned} \rho(\mathbf{r}) &= -\frac{1}{\pi}\int_{-\infty}^{W_f} \mathrm{d}W \,\mathrm{Im}\,\mathrm{Tr}\, G^+(\mathbf{r},\mathbf{r};W+\mathrm{i}\eta) \\ n(W) &= -\frac{1}{\pi}\int_{\mathbb{V}} \mathrm{d}^3\mathbf{r} \,\mathrm{Im}\,\mathrm{Tr}\, G^+(\mathbf{r},\mathbf{r};W+\mathrm{i}\eta) \end{aligned} \tag{2.97}$$

These expressions differ from the non-relativistic case by involving the trace of the Green's function. The trace of a matrix is the sum of the diagonal elements and is included because of 4-component nature of the wavefunction:

$$|\psi|^2 = \sum_{i=1}^{4} \psi_i^\dagger \psi_i = \mathrm{Tr}\,\psi \otimes \psi^\dagger. \tag{2.98}$$

The spectral representation is also useful because it enables us to easily examine the poles of the Green's function which will be useful in the investigation of the embedding method later on. To conclude this section on the properties of Green's function we will show a practical way to construct them by the direct method.

2.2.2 The Direct Method for the Green's Function

So far we have presented two methods for determining the relativistic Green's function. We can solve the Dirac equation directly for the Green's function (2.94) or use the eigenvalues and eigenvectors of H and the spectral representation. These are not always the most practical methods, for example if the eigenvalue spectrum of H is continuous the spectral representation requires a contour integral in the complex energy plane. It is often easier to use the *direct method*. This is particularly applicable to one-dimensional systems or where we can define an ordering of coordinates. For example we can define $r < r'$ or $r > r'$ in spherical polar coordinates $\mathbf{r} = (r, \theta, \phi)$, or for a surface in the $x - y$ plane we can separate out the parallel and perpendicular directions $\mathbf{r} = (\mathbf{r}_{||}, z)$ and put $z > z'$ or $z < z'$.

For simplicity we will consider a one dimensional system at real energies in the absence of a vector potential. The wavefunction, $\psi(z)$, satisfies the Dirac equation

$$-\mathrm{i}c\hbar\alpha_z \frac{\mathrm{d}}{\mathrm{d}z}\psi(z) + \left[\beta mc^2 + V(z) - W\right]\psi(z) = 0 \tag{2.99}$$

where $V(z)$ is also taken to be real. The Hermitian conjugate of this equation is

$$\mathrm{i}c\hbar \frac{\mathrm{d}}{\mathrm{d}z}\psi^\dagger(z)\alpha_z + \psi^\dagger(z)\left[\beta mc^2 + V(z) - W\right] = 0. \tag{2.100}$$

Multiplying (2.99) from the left by $\psi^\dagger(z)$, (2.100) from the right by $\psi(z)$ and subtracting we have

$$\mathrm{i}c\hbar\frac{\mathrm{d}}{\mathrm{d}z}\left[\psi^\dagger(z)\alpha_z\psi(z)\right]=0 \tag{2.101}$$

which implies

$$\psi^\dagger(z)\alpha_z\psi(z)=\lambda \tag{2.102}$$

where λ is a constant. The equation satisfied by the Green's function corresponding to (2.99) is

$$-\mathrm{i}c\hbar\alpha_z\frac{\mathrm{d}}{\mathrm{d}z}G(z,z';W)+\left[\beta mc^2+V(z)-W\right]G(z,z';W)=-\delta(z-z') \tag{2.103}$$

For $z\neq z'$ we can write

$$\begin{aligned}G_>(z,z';W)&=\psi_>(z)A(z'),\qquad z>z'\\ G_<(z,z';W)&=\psi_<(z)B(z'),\qquad z<z'\end{aligned} \tag{2.104}$$

where $A(z')$ and $B(z')$ are some 1×4 spinor functions of z' and an outer product is implied. We require that the wavefunction for $z<z'$, $\psi_<(z)$ is well behaved at $z=-\infty$ and the wavefunction for $z>z'$, $\psi_>(z)$ is well behaved at $z=+\infty$, i.e. outgoing boundary conditions.

To find $A(z)$ and $B(z)$ we note that integrating across the delta function in the equation satisfied by the Green's function

$$-\mathrm{i}c\hbar\alpha_z\int_{z-\epsilon}^{z+\epsilon}\mathrm{d}z\,\frac{\mathrm{d}}{\mathrm{d}z}G(z,z';W)+\int_{z-\epsilon}^{z+\epsilon}\mathrm{d}z\,\left[\beta mc^2+V(z)-W\right]G(z,z';W)=-\int_{z-\epsilon}^{z+\epsilon}\mathrm{d}z\,\delta(z-z') \tag{2.105}$$

and letting $\epsilon\to 0$ we get

$$-\mathrm{i}c\hbar\left[G_>(z,z;W)-G_<(z,z;W)\right]=\mathbb{I}_4, \tag{2.106}$$

where $\mathbb{I}_4$ is the 4×4 unit matrix. Substituting in from (2.104) gives

$$\alpha_z\psi_>(z)A(z)-\alpha_z\psi_<(z)B(z)=\frac{\mathrm{i}}{c\hbar}\mathbb{I}_4. \tag{2.107}$$

Next we multiply from the left by $\psi_>^\dagger(z)$ and use (2.102) to get

$$\lambda_{>>}A(z)-\lambda_{><}B(z)=\frac{\mathrm{i}}{c\hbar}\psi_>^\dagger(z) \tag{2.108}$$

and similarly by multiplying (2.107) from the left by $\psi^\dagger_<(z)$

$$\lambda_{<>}A(z) - \lambda_{<<}B(z) = \frac{\mathrm{i}}{c\hbar}\psi^\dagger_<(z) \tag{2.109}$$

where

$$\lambda_{>>} = \psi^\dagger_>(z)\alpha_z\psi_>(z) \qquad \text{etc.} \tag{2.110}$$

Rearranging gives expressions for $A(z)$ and $B(z)$:

$$A(z) = \frac{\mathrm{i}}{c\hbar(\lambda_{><}\lambda_{<>} - \lambda_{<<}\lambda_{>>})}\left[\lambda_{><}\psi^\dagger_<(z) - \lambda_{<<}\psi^\dagger_>(z)\right]$$

$$B(z) = \frac{\mathrm{i}}{c\hbar(\lambda_{><}\lambda_{<>} - \lambda_{<<}\lambda_{>>})}\left[\lambda_{>>}\psi^\dagger_<(z) - \lambda_{<>}\psi^\dagger_>(z)\right], \tag{2.111}$$

which, along with (2.104) allows us to construct the Green's function for our one-dimensional problem which satisfies outgoing boundary conditions.

As an example we will use the direct method for the one-dimensional free particle Green's function. The electron (positive energy) wavefunctions satisfying the appropriate boundary conditions at $\pm\infty$ are

$$\psi_>(z) = \begin{pmatrix} \phi_{m_s} \\ \gamma\sigma_z k_z \phi_{m_s} \end{pmatrix} \mathrm{e}^{\mathrm{i}kz}, \qquad \psi_<(z) = \begin{pmatrix} \phi_{m_s} \\ -\gamma\sigma_z k_z \phi_{m_s} \end{pmatrix} \mathrm{e}^{-\mathrm{i}kz} \tag{2.112}$$

where $\gamma = c\hbar/(W + mc^2)$ with the energy given by $W = \sqrt{c^2\hbar^2k^2 - m^2c^4}$. First of all we calculate the $\lambda_{>>}$ etc from (2.102);

$$\lambda_{>>} = -\lambda_{<<} = 2\gamma k\phi^\dagger_{m_s}\phi_{m'_s}, \qquad \lambda_{<>} = \lambda_{><} = 0 \tag{2.113}$$

and substitute these into the expressions for $A(z)$ and $B(z)$ to get

$$A(z) = \frac{\mathrm{i}}{2c\hbar\gamma k}\psi^\dagger_>(z), \qquad B(z) = \frac{\mathrm{i}}{2c\hbar\gamma k}\psi^\dagger_<(z). \tag{2.114}$$

Finally, using (2.104) we obtain the one-dimensional, relativistic free particle Green's function

$$G(z,z';W) = \frac{-\mathrm{i}}{2c\hbar\gamma k}\begin{pmatrix} \mathbb{I}_2 & \gamma k\,\mathrm{sgn}(z-z')\sigma_z \\ \gamma k\,\mathrm{sgn}(z-z')\sigma_z & \gamma^2k^2\mathbb{I}_2 \end{pmatrix} \mathrm{e}^{\mathrm{i}k|z-z'|} \tag{2.115}$$

Now that we have an expression for the relativistic Green's function it is worth investigating its non-relativistic limit. As $c \to \infty$,

$$c\hbar\gamma = \frac{c^2\hbar^2}{W + mc^2} \to \frac{\hbar^2}{2m}, \tag{2.116}$$

where we have separated out the rest energy by putting $W = E + mc^2$. The limit is

$$G(z, z'; E) = -\frac{2m}{\hbar^2}\frac{\mathrm{i}}{2k}\mathrm{e}^{\mathrm{i}k|z-z'|}\begin{bmatrix} \mathbb{I}_2 & 0 \\ 0 & 0 \end{bmatrix}. \tag{2.117}$$

This is exactly the non-relativistic free particle Green's function in one-dimension [41]. There are some important things to notice. The first is that $G(z, z'; E)$ is diagonal in spin as we would expect. Secondly, only the large-large component has survived in the non-relativistic limit, which is not surprising since we have already seen that the large component of the wavefunction reduces to the non-relativistic wavefunction in the $c \to \infty$ limit.

The direct method will be useful in later chapters for constructing the Green's function for simple systems used to investigate the properties and non-relativistic limits of the transfer matrix and expressions for transport problems.

Chapter 3

ELECTRONIC STRUCTURE CALCULATIONS

In the previous chapter we introduced the Dirac equation and discussed some of its properties and looked in detail at solutions for free-particles and spherically symmetric systems, but we weren't particularly concerned with the effects of having more than one electron. In atoms, molecules and solids we are really dealing with many electrons which all interact with each other as well as with any external potentials and the situation becomes complicated very quickly. In this chapter we introduce Density functional Theory (DFT) which enables systems with many electrons to be treated within a single-particle framework. We begin with a presentation of non-relativistic DFT, discussing in detail the theoretical background and some of the assumptions and approximations necessary to apply it to realistic problems. Finally we outline the extension to relativistic problems.

Quantum mechanics tells us that the simple equation,

$$H\Psi = E\Psi, \tag{3.1}$$

allows the wavefunction, Ψ, of the system to be calculated, which contains all the information about its physical properties. To construct the Hamiltonian H, we need to know the scalar and vector potentials acting on the particles in the system. The problems begin when we consider that even a single mole of substance contains around 10^{23} atoms with charged nuclei and electrons that all interact with each other. Given that the positions of each nucleus and electron each require 3 spatial coordinates means even storing these coordinates in computer memory is impossible. It is clear then, that some clever tricks are needed.

Perhaps the most widely used approach to dealing with the complicated many-body interactions in solid state physics is *Density Functional Theory* (DFT). There are two main ideas in DFT. The first is that we don't actually need to calculate

the many-body wavefunction as the fundamental quantity, but simply the electronic charge *density*. The justification for this is supplied by the Hohenberg-Kohn theorems and the introduction of energy *functionals*. The second idea, due to Kohn and Sham, leads to a system of single particle equations that allow the density to be calculated. DFT therefore reduces the complex many-body problem, requiring 3×10^{23} coordinates, to a single particle theory based on the density which depends on only 3 coordinates. The foundations of DFT, the Hohenberg-Kohn theorem and Kohn-Sham equations, are exact but in practice they can only be implemented by making certain approximations concerning the electron-electron interactions. Much of the continuing development of DFT is concerned with improvements in these approximations.

DFT is particularly well suited to calculating ground state properties of molecules and solids; bulk moduli, lattice constants, band structures and densities of states for a large range of materials are accurately reproduced. Nevertheless there are systems where DFT fails, in particular it gives poor estimates of the band gap in some semiconductors and is insufficient for describing strongly correlated materials such as many transition metal oxides. There are a number of extensions and alternatives to DFT which attempt to address these problems, for example the GW method and Time-Dependent DFT. However, in this work we are concerned primarily with ground state properties of metals and DFT is a perfectly suitable framework in which to incorporate the embedding method for these materials.

3.1 Density Functional Theory

As we have already mentioned, DFT has a central role in the majority of electronic structure calculations and seeks to simplify the problem of a large number of interacting electrons. The Born-Oppenheimer approximation separates the motion of the electrons and nuclei and to a certain extent simplifies the problem we face when trying to calculate total wavefunctions. However, we still have 10^{23} interacting electrons which is too many to realistically deal with. DFT solves this problem by basically turning it on its head. The total wavefunction can be used to calculate all the physical properties of a system. One of these properties is the density of electrons and the fundamental principle of DFT is that we can replace the central role of the wavefunction with the electron density, turning a many electron problem into a single electron problem. Starting from an initial 'guess' for the electron density, via an iterative procedure which involves the construction of *single* particle Hamiltonians we can converge to the correct density. From the correct density we can construct an effective potential in which the electrons move and use this to calculate single-particle wavefunctions and Green's functions describing the many particle system.

3.1.1 The Born-Oppenheimer Approximation

For a system of interacting electrons and nuclei we can write the Hamiltonian in (3.1) as

$$H = T_n + V_{n-n} + T_e + V_{e-e} + V_{e-n}, \tag{3.2}$$

where T_n is the kinetic energy of the nuclei, V_n describes the nucleus-nucleus interaction, T_e is the kinetic energy of the electrons, V_{e-e} the electron-electron interaction and V_{e-n} the interaction between electrons and nuclei.

The total wavefunction for the k^{th} eigenstate of the system satisfies (3.1)

$$H\Psi_k(\mathbf{r}, \mathbf{R}) = E_k\Psi_k(\mathbf{r}, \mathbf{R}), \tag{3.3}$$

where $\mathbf{r}$ are the electron coordinates and $\mathbf{R}$ the nuclear coordinates. For a fixed, instantaneous configuration of nuclei, $[\mathbf{R}]$, the electron eigenfunctions, $\psi_i(\mathbf{r} : [\mathbf{R}])$ form a complete set of (normalised) states with eigenvalues $E_i([\mathbf{R}])$ at each $\mathbf{R}$. These electron eigenfunctions depend on the nuclear coordinates and we can write the full wavefunction as

$$\Psi_k(\mathbf{r}, \mathbf{R}) = \sum_i \chi_{ik}([\mathbf{R}])\psi_i(\mathbf{r} : [\mathbf{R}]) \tag{3.4}$$

and the eigenstates of the complete electron-nuclear system are now specified by the expansion coefficients $\chi_{ki}([\mathbf{R}])$ which are functions of the nuclear coordinates. Substituting (3.4) into (3.3), multiplying from the left by $\psi_j^*(\mathbf{r} : [\mathbf{R}])$ and integrating over the electron coordinates, $\mathbf{r}$, we have

$$\left[T_n + E_i([\mathbf{R}]) - E_k\right]\chi_{ki}([\mathbf{R}]) = -\sum_j U_{ij}\chi_{kj}([\mathbf{R}]), \tag{3.5}$$

where the matrix element U_{ij} contains the electron-nuclear interaction. The mass of the nucleus is typically of the order of 10^4 times the mass of the electron. As a result, electrons respond much more rapidly to motion of the nuclei than the nuclei react to changes in the electron positions. In other words, to a good approximation, changes in $\mathbf{R}$ do not cause transitions between electron eigenstates and the off diagonal terms, $i \neq j$ in U_{ij} can be ignored. This is the *Born–Oppenheimer approximation* [49] within which, although the electron eigenstates and eigenvalues of a given state change as $\mathbf{R}$ changes, there are no excitations, or changes from one electron state to another, induced by the nuclear motion; the interaction is *adiabatic* [30]. The nuclear wavefunction can be calculated from (3.5) where the potential and kinetic energy of the electrons, $E_i([\mathbf{R}])$, acts as an effective potential [45].

The main advantage of the Born-Oppenheimer approximation is that it decouples the motion of the electrons and nuclei for a fixed arrangement of nuclei. This is a valid assumption unless we are concerned with electron excitations induced by motions of the crystal lattice, i.e. electron-phonon interactions. Superconductors

are an example of a class of systems where electron-phonon interactions are crucial to understanding physical properties. However, in this work we are concerned with ground state electronic properties of metals and the Born-Oppenheimer approximation is a valid method of decoupling the electrons and nuclei and we may now focus on the electrons in our system.

3.1.2 The Hohenberg–Kohn Theorems

Density Functional Theory takes as its starting point the density, $n(\mathbf{r})$, of the electrons in the system. For N electrons with normalised many–body wavefunction $\psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)$ the density is

$$n(\mathbf{r}) = \int \mathrm{d}^3\mathbf{r}_1 \cdots \int \mathrm{d}^3\mathbf{r}_N \, \delta(\mathbf{r} - \mathbf{r}_n)\Psi^*(\mathbf{r}_1, \cdots, \mathbf{r}_N)\Psi(\mathbf{r}_1, \cdots, \mathbf{r}_N) \tag{3.6}$$

and we assume that the number of electrons is fixed. The electrons interact with each other via the Coulomb potential and feel some, as yet unspecified, external potential, V_{ext}. The only variable here is the external potential, therefore it is this which determines the ground state wavefunction and hence the ground state density, n_0. Hohenberg and Kohn [50] showed that the opposite of this is also true; that the external potential can be calculated from the ground state density. The first Hohenberg-Kohn theorem is

> *For a system of interacting particles in an external potential V_{ext}, this potential is uniquely determined (up to a constant) by the ground state particle density $n_0(\mathbf{r})$.*

A corollary of this is that since the things we need to construct the Hamiltonian (namely, V_{ext}) is determined by the ground state density, then the many full body wavefunction is also given by n_0. Therefore *all* of the properties of the system come from n_0. The density, n_0, only depends on 3 spatial coordinates, rather than the $3N$ necessary for a complete description of the wavefunction, in principle greatly simplifying the problem.

The next question is how to extract physical properties from n_0. To do this we define a total energy *functional*. A functional is an operation (often via an integral) which maps a function onto a scalar. For example the potential, ϕ, due to a charge distribution $\rho(\mathbf{r})$, is a functional given by

$$\phi\left[\rho(\mathbf{r})\right] = \frac{1}{4\pi\epsilon_0}\int \mathrm{d}^3\mathbf{r}' \, \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}. \tag{3.7}$$

If the electron density, n, via the first Hohenberg-Kohn theorem, is connected to the wavefunction ψ, then the total energy functional is

$$E[n(\mathbf{r})] = \langle\psi|H|\psi\rangle = \int \mathrm{d}^3\mathbf{r} n(\mathbf{r})V_{ext}(\mathbf{r}) + F[n(\mathbf{r})], \tag{3.8}$$

where the functional $F[n(\mathbf{r})]$ contains the classical coulomb electron-electron interactions, nuclei-nuclei interactions and the electron kinetic energy. If we put all of the difficult many-body effects, electron-nuclei interactions and external potentials into $V_{ext}(\mathbf{r})$, then $F[n(\mathbf{r})]$ is *universal*, i.e. the same for all systems.

The second Hohenberg Kohn theorem shows us how to get the ground state density from $E[n(\mathbf{r})]$:

> *For a given density, we can construct a universal energy functional. For a particular external potential, the exact ground state energy of the system is the global minimum value of this functional. The density that minimises the functional is the exact ground state density.*

In other words, if we can construct the energy functional $E[n(\mathbf{r})]$ and find its global minimum (subject to the constraint that the number of particles remains fixed), then we can extract the ground state density from this minimum functional.

So far, the Hohenberg-Kohn theorems have simply been stated and, since it is relatively straightforward, it is worth justifying them. Regarding the first theorem, suppose that there are two *different* external potentials, $V_{ext}^{(1)}(\mathbf{r})$ and $V_{ext}^{(2)}(\mathbf{r})$, which give the same ground state density, $n_0(\mathbf{r})$. These two potentials are associated with Hamiltonians, $H^{(1)}$ and $H^{(2)}$ with ground state wavefunctions $\psi^{(1)}$ and $\psi^{(2)}$. The wavefunction $\psi^{(2)}$ is not the ground state of $H^{(1)}$, therefore

$$E^{(1)} = \langle\psi^{(1)}|H^{(1)}|\psi^{(1)}\rangle < \langle\psi^{(2)}|H^{(1)}|\psi^{(2)}\rangle. \tag{3.9}$$

where the strict inequality is because we assume non-degenerate ground states. The last term can be written (remembering that $F[n(\mathbf{r})]$ in (3.8) is universal)

$$\begin{aligned}\langle\psi^{(2)}|H^{(1)}|\psi^{(2)}\rangle &= \langle\psi^{(2)}|H^{(2)}|\psi^{(2)}\rangle + \langle\psi^{(2)}|H^{(1)} - H^{(2)}|\psi^{(2)}\rangle \\ &= E^{(2)} + \int \mathrm{d}^3\mathbf{r}\, \left[V_{ext}^{(1)}(\mathbf{r}) - V_{ext}^{(2)}(\mathbf{r})\right] n_0(\mathbf{r})\end{aligned} \tag{3.10}$$

and it follows that

$$E^{(1)} < E^{(2)} + \int \mathrm{d}^3\mathbf{r}\, \left[V_{ext}^{(1)}(\mathbf{r}) - V_{ext}^{(2)}(\mathbf{r})\right] n_0(\mathbf{r}). \tag{3.11}$$

Doing the same thing for $E^{(2)}$ leads to

$$E^{(2)} < E^{(1)} + \int \mathrm{d}^3\mathbf{r}\, \left[V_{ext}^{(2)}(\mathbf{r}) - V_{ext}^{(1)}(\mathbf{r})\right] n_0(\mathbf{r}) \tag{3.12}$$

and if we add these together we get

$$E^{(1)} + E^{(2)} < E^{(2)} + E^{(1)}, \tag{3.13}$$

which is a clear contradiction. We cannot therefore have two different external potentials which give rise to the same ground state density.

To justify the second Hohenberg-Kohn theorem we consider a ground state density, $n_0(\mathbf{r})$, which corresponds to an external potential $V_{ext}^{(0)}(\mathbf{r})$. The ground state energy, E_0, equal to the total energy functional, $E_0[n_0(\mathbf{r})]$, is the expectation value of the Hamiltonian, $H^{(0)}$ derived from the density $n_0(\mathbf{r})$:

$$E_0 = E_0[n_0(\mathbf{r})] = \langle \psi^{(0)} | H^{(0)} | \psi^{(0)} \rangle. \tag{3.14}$$

A different energy, E_1, (not the ground state) corresponds to a different wavefunction, $\psi^{(1)}$. Because it is not the ground state, E_1 is greater than E_0 and

$$\langle \psi^{(0)} | H^{(0)} | \psi^{(0)} \rangle \; < \; \langle \psi^{(1)} | H^{(0)} | \psi^{(1)} \rangle \tag{3.15}$$

hence

$$E_0 < E_1. \tag{3.16}$$

The energy corresponding to the ground state is therefore less than any other functional corresponding to a different density to within a constant. In other words, finding the minimum of the total energy functional with respect to variations in the density gives the ground state density.

All of this is very insightful, having reduced the problem of calculating the many body wavefunction to finding the ground state density, but it provides no method by which this can be achieved in practice.

3.1.3 The Kohn–Sham Equations

From the Hohenberg-Kohn theorems, we know that we need to find the ground state electron density, $n_0(\mathbf{r})$, which we do by minimising the functional $E[n(\mathbf{r})]$ with respect to the density, subject to the constraint that the number of particles is fixed. In principle, the Hohenberg-Kohn theorems make no approximations and are exact but they provide no insight into how we practically calculate the density. The difficulties arise from the fact that the system is described by a many-body wavefunction since all the interactions do not allow single particle eigenstates, making construction of the density virtually impossible.

The Kohn-Sham *ansatz* [51] assumes that, given some interacting system of electrons and nuclei with density $n(\mathbf{r})$, there exists a *non-interacting* system with the same density. Since the density is the fundamental quantity we require, we can look for the density of the auxiliary, non-interacting system. The advantage of a non-interacting system is that we can solve it using single particle wavefunctions,

$\psi_i(\mathbf{r})$, meaning that the density, for N electrons, is

$$n(\mathbf{r}) = \sum_{i=1}^{N} |\psi_i(\mathbf{r})|^2 . \tag{3.17}$$

According to the Kohn-Sham scheme

$$\tilde{T}[n(\mathbf{r})] = -\frac{\hbar^2}{2m} \sum_{i=1}^{N} \langle \psi_i | \nabla^2 | \psi_i \rangle \tag{3.18}$$

The energy functional of the interacting system can be written

$$E[n(\mathbf{r})] = \tilde{T}[n(\mathbf{r})] + E_{ext}[n(\mathbf{r})] + E_H[n(\mathbf{r})] + E_{e-e}[n(\mathbf{r})]. \tag{3.19}$$

The first term is the kinetic energy of a system of non-interacting electrons and the second term describes any external potentials and the interaction between the electrons and the electrostatic potential of the nuclei. The third term, called the *Hartree term*, is the classical electron-electron Coulomb interaction:

$$E_H[n(\mathbf{r})] = e^2 \int \mathrm{d}^3\mathbf{r} \int \mathrm{d}^3\mathbf{r}' \, \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} = \int \mathrm{d}^3\mathbf{r} \, n(\mathbf{r}) V_H(\mathbf{r}) \tag{3.20}$$

where $V_H(\mathbf{r})$ is the Hartree potential. The final term contains all of the many-body interactions beyond the Hartree term, including corrections to the kinetic energy.

The Kohn-Sham ansatz allows us to find the density which minimises (3.19) within a single-particle formulation. In doing this we make use of an effective equation which gives single-particle eigenstates and eigenvalues and use these to construct the density via (3.17). Before going into the details of this, it should be mentioned that these wavefunctions and eigenvalues do not, in general, correspond to real electron states and energies of the real, fully interacting system but are merely a convenient stage in finding the true ground state density.

To derive the Kohn-Sham equations we minimise (3.19) with respect to the density $n(\mathbf{r})$ subject to the constraint that the number of particles, N, is fixed. The density can be written in terms of the single-particle Kohn-Sham orbitals, $|\psi_i\rangle$,

$$n(\mathbf{r}) = \sum_{i=1} |\psi_i(\mathbf{r})|^2 , \tag{3.21}$$

where the sum is over the lowest N energy eigenstates. The constraint that the number of particles, N, is fixed is incorporated by introducing the Lagrange multiplier ϵ, such that

$$\int \mathrm{d}^3\mathbf{r} \, n(\mathbf{r}) = N, \tag{3.22}$$

and[‡]

$$\frac{\delta}{\delta n(\mathbf{r})}\left[E[n(\mathbf{r})] - \epsilon \int \mathrm{d}^3\mathbf{r}\, n(\mathbf{r})\right] = 0 \tag{3.23}$$

Using the chain rule to minimise $E[n(\mathbf{r})]$ with respect to variations in $\langle\psi_i|$:

$$\frac{\delta E}{\delta \psi_i^*(\mathbf{r})} = \frac{\delta \tilde{T}}{\delta \psi_i^*(\mathbf{r})} + \left[\frac{\delta E_{ext}}{\delta n(\mathbf{r})} + \frac{E_H}{\delta n(\mathbf{r})} + \frac{\delta E_{e-e}}{\delta n(\mathbf{r})}\right]\frac{\delta n(\mathbf{r})}{\delta \psi_i^*(\mathbf{r})} = \epsilon_i \psi_i(\mathbf{r}). \tag{3.24}$$

From the definitions of the kinetic energy, $\tilde{T}$, and the density in terms of the orbitals:

$$\frac{\delta \tilde{T}}{\delta \psi_i^*(\mathbf{r})} = -\frac{\hbar^2}{2m}\nabla^2\psi_i(\mathbf{r}), \qquad \frac{\delta n(\mathbf{r})}{\delta \psi_i^*(\mathbf{r})} = \psi_i(\mathbf{r}), \tag{3.25}$$

we arrive at an effective Schrödinger equation for the Kohn-Sham orbitals, $\psi_i(\mathbf{r})$ and 'energies', ϵ_i:

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V_{eff}(\mathbf{r})\right]\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}). \tag{3.26}$$

The effective potential is

$$V_{eff}(\mathbf{r}) = V_{ext}(\mathbf{r}) + V_H(\mathbf{r}) + V_{e-e}(\mathbf{r}), \tag{3.27}$$

where

$$V_{ext}(\mathbf{r}) = \frac{\delta E_{ext}[n(\mathbf{r})]}{\delta n(\mathbf{r})} \quad \text{or} \quad E_{ext}[n(\mathbf{r})] = \int \mathrm{d}^3\mathbf{r}\, n(\mathbf{r}) V_{ext}(\mathbf{r}) \tag{3.28}$$

and

$$V_{e-e}(\mathbf{r}) = \frac{\delta E_{e-e}[n(\mathbf{r})]}{\delta n(\mathbf{r})} \quad \text{or} \quad E_{e-e}[n(\mathbf{r})] = \int \mathrm{d}^3\mathbf{r}\, n(\mathbf{r}) V_{e-e}(\mathbf{r}). \tag{3.29}$$

Equations (3.26)-(3.29) are the *Kohn-Sham equations*, solved to find the single particle eigenstates from which we construct the density (3.21). The total energy functional can be written

$$E[n(\mathbf{r})] = \sum_{i=1}^{N}\epsilon_i - \frac{1}{2}\int \mathrm{d}^3\mathbf{r}\int \mathrm{d}^3\mathbf{r}'\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} - \int \mathrm{d}^3\mathbf{r} n(\mathbf{r})\frac{\delta E_{\mathrm{xc}}[n(\mathbf{r})]}{\delta n(\mathbf{r})} + E_{\mathrm{xc}}[n(\mathbf{r})]. \tag{3.30}$$

The Hohenberg-Kohn theorems tell us that we need to find the electron density which minimises the total energy functional of our system and the Kohn-Sham equations provide a way of constructing the density from single particle wavefunctions, assuming that we know the necessary effective potential, which itself depends on the density. It is therefore necessary for the Kohn-Sham equations to be solved itera-

[‡]The following expressions contain functional derivatives which are defined in a similar way to standard derivatives. For a functional $F[\phi(\mathbf{r})]$

$$\frac{\delta F[\phi(\mathbf{r})]}{\delta \phi(\mathbf{r})} = \lim_{\mu\to 0}\frac{F[\phi(\mathbf{r}) + \mu\delta(\mathbf{r}-\mathbf{r}')] - F[\phi(\mathbf{r})}{\mu}$$

tively until the output density is the same as the input density; self-consistency. An idealised self-consistency procedure is shown in fig 3-1. To begin the cycle, an initial charge density, $n(\mathbf{r})$, is constructed. This is typically a superposition of atomic charge densities. In the next step, the effective potential is calculated. The Hartree contribution to the potential is typically calculated by solving the Poisson equation,

$$\nabla^2 V_H(\mathbf{r}) = -n(\mathbf{r}), \tag{3.31}$$

the details of this and the V_{e-e} term being specific to the particular type of system studied. This is considered in more detail in later chapters. Once the effective potential has been generated, it is inserted into the Kohn-Sham Hamiltonian and the eigenvalues and eigenstates are calculated. A new charge density, $n'(\mathbf{r})$ is then constructed. The next step is to compare the difference in the original charge density, $n(\mathbf{r})$ and the new one, $n'(\mathbf{r})$. A parameter, η, is usually chosen such that if $|n'(\mathbf{r}) - n(\mathbf{r})| < \eta$ the density is considered converged and the calculation is done. If $|n'(\mathbf{r}) - n(\mathbf{r})| > \eta$, then we haven't found the ground state density and we must restart the cycle. Rather than simply using the new density as an input density for the next cycle, it is both more accurate and efficient to mix the old and new densities. There are a number of schemes for doing this from simple linear mixing [45] (shown in fig 3-1) to the more efficient Anderson [52] and Broyden [53], [54] methods.

As far as it has been summarised here, the only simplification we have made is the Born-Oppenheimer approximation and to within this limit DFT is exact. The Hohenberg-Kohn theorems simply justify the use of the ground state electron density as a replacement for the many-body wavefunction and the Kohn-Sham equations provide a mechanism by which we can practically calculate the density within a single-particle framework. In principle if we have exact expressions for all of the terms in the total energy functional we have a complete method. The Hartree term describes the correlation of the electrons due to the classical Coulomb interaction; how each electron interacts with the charge distribution of all the other electrons. However, we have hidden the additional quantum mechanical electron-electron interactions in the potential V_{e-e}. These additional interactions are commonly referred to as the exchange and correlation interactions and arise from the Pauli exclusion principle and correlation effects beyond the Hartree term. The potential, V_{e-e} is usually referred to as the *exchange-correlation potential* and is denoted from now on by V_{xc}. The full exchange and correlation effects can only be exactly determined for the simplest systems and the following section describes some of the commonest approximations used in DFT calculations.

3.1.4 The Exchange–Correlation Potential

The accuracy of DFT calculations depends on the accuracy of the exchange correlation potential. Unfortunately it is extremely difficult to calculate the exchange-

correlation effects for all but the simplest systems. However, the exchange correlation potential is universal and any suitable approximation will be applicable to all systems, with, of course, varying degrees of accuracy. In this section, we will discuss the widely used and surprisingly successful Local Density Approximation (LDA) and its extension via Generalised Gradient Approximations (GGA).

To illustrate the interactions described by the exchange-correlation potential it is useful to define a pair correlation function [10]

$$P(\mathbf{r}_1, \mathbf{r}_2) = \langle \psi | \hat{n}(\mathbf{r}_1)\hat{n}(\mathbf{r}_2) | \psi \rangle, \tag{3.32}$$

where $\hat{n}(\mathbf{r}_i)$ is an operator which measures the electron density at $\mathbf{r}_i$. The Pauli exclusion principle requires that the spatial part of the many-electron wavefunction changes sign when two particles are exchanged. This means that two particles cannot be at the same position[‡] and $P(\mathbf{r}_1, \mathbf{r}_2) \to 0$ as $\mathbf{r}_1 \to \mathbf{r}_2$. As $\mathbf{r}_1 - \mathbf{r}_2 \to \infty$ there should be no interaction between electrons and $P(\mathbf{r}_1, \mathbf{r}_2) \to n(\mathbf{r}_1)n(\mathbf{r}_2)$, the product of the densities. The main point of this is that electrons tend to avoid each other not just via the Coulomb repulsion due to their charge but because of the quantum mechanical effects of the Pauli principle, or exchange. A common way of interpreting these effects is by the *exchange-correlation hole* [55], [45], an area of decreased density around each electron. This 'hole' leads to a reduced interaction between electrons, resulting in a screening effect. The exchange-correlation energy can therefore be written [45], [10]

$$E_{\mathrm{xc}}[n(\mathbf{r})] = \frac{1}{2} \int \mathrm{d}^3\mathbf{r} \int \mathrm{d}^3\mathbf{r}' \, \frac{n(\mathbf{r})\tilde{n}_{\mathrm{xc}}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}. \tag{3.33}$$

The factor of 1/2 ensures that interactions are not double counted and $\tilde{n}_{\mathrm{xc}}(\mathbf{r}, \mathbf{r}')$ represents the interaction of an electron with its own exchange-correlation hole:

$$\tilde{n}_{\mathrm{xc}}(\mathbf{r}, \mathbf{r}') = \int_0^1 \mathrm{d}\lambda \, n_{\mathrm{xc}}^{\lambda}(\mathbf{r}, \mathbf{r}'), \tag{3.34}$$

where λ is a coupling constant for the interaction and $n_{\mathrm{xc}}^{\lambda}(\mathbf{r}, \mathbf{r}')$ is the depletion (hole) in the density of an electron at $\mathbf{r}'$ due to an electron at $\mathbf{r}$.

In general, the exchange correlation energy is not known exactly, but there is a special case where it is relatively straightforward to calculate it very accurately. For a system under the influence of a potential which is independent of position, the density is constant and we have a homogeneous electron gas. The exchange energy of a homogeneous electron gas is known in an exact analytical form, but the correlation

[‡]The Pauli principle states

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N) = -\Psi(\mathbf{r}_2, \mathbf{r}_1, \cdots, \mathbf{r}_N)$$

hence $\Psi = 0$ if $\mathbf{r}_1 = \mathbf{r}_2$.

effects must be obtained numerically [56], Quantum Monte Carlo methods [57] being commonly used to generate parameterizations of the correlation effects. The most widely used parameterizations are those of Perdew and Zunger [58], Perdew and Wang [59] and Vosko, Wilkes and Nusair [60].

Kohn and Sham suggested [51] that many solids can reasonably be considered very close to a homogeneous electron gas and the effects of exchange and correlation can be assumed to be local. This is the *Local Density Approximation* (LDA). The assumption is that

$$E_{\mathrm{xc}}[n(\mathbf{r})] = \int \mathrm{d}^3\mathbf{r}\, n(\mathbf{r}) \epsilon_{\mathrm{xc}}^{\mathrm{HEG}}(n(\mathbf{r})). \tag{3.35}$$

In the LDA we have essentially divided the system into small volume elements in which the density is constant and used the exchange-correlation energy of the homogeneous electron gas with the appropriate density in each volume. Note that the exchange-correlation energy of the homogeneous gas is a *function* of the density rather than a functional because of the local nature of the exchange-correlation effects. The LDA can be extended to include spin via the Local Spin-Density Approximation (LSDA) [61].

Considering that it is based on the assumption that the density varies gradually with position, the LDA is surprisingly successful for calculating many physical properties. A particular reason for this is that the exchange-correlation hole obeys the correct sum rule that the charge displaced by the hole is exactly one electron [‡] [61]. Additionally, the shape of the hole does not need to be known exactly since in the general case the exchange-correlation energy only depends on the hole averaged over a sphere [61] [62].

Despite its many successes, the LDA fails to accurately reproduce experimental results in some instances. A particularly well known case is the band gap in semiconductors [63]. For example, the band gap in silicon is 50% too small and in germanium 100% too small, with the valence and conduction bands predicted to overlap [64]. The LDA also underestimates binding energies for surface adsorption [65]. Due to the local assumption in the LDA it performs particularly poorly for systems where the electron-electron correlations are highly non-local [66]. This is particularly true of several transition metal oxides and the LDA predicts that the ground state of NiO is metallic when it is in fact an insulator [67]. Transition metal oxides are particularly interesting because they include a number of high-temperature superconductors.

The LDA assumes that the density is slowly varying and a first approximation to extending this to include spatial variations to the density are Generalised Gradient Approximations (GGA) [62]. These include terms which vary as the magnitude of the gradient of the density and are an adaptation of the Gradient expansion Approximation (GEA) [68]. Modifications to the GEA are necessary since it violates the

[‡]This is to be expected because the hole exists to ensure that at each position there is only one electron, not two. In other words, the electron at $\mathbf{r}$ has 'pushed away' another electron.

sum rules and the independence of the exchange-correlation potential on the shape of the exchange-correlation hole, which are preserved in the LDA, and gradients in real materials are large enough to cause the expansions to break down [62]. GGAs improve the GEA by providing functionals which are well behaved at large gradients. Functionals involving the gradient of the density have the general form

$$E_{\rm xc}^{GGA}[n(\mathbf{r})] = \int \mathrm{d}^3\mathbf{r}\, n(\mathbf{r})\epsilon_{\rm x}^{\rm heg}(n)F_{\rm xc}(n, |\nabla n|, |\nabla n|^2, \cdots) \tag{3.36}$$

where $\epsilon_{\rm x}^{\rm heg}(n)$ is the exchange energy of the homogeneous electron gas, which can be calculated analytically. $F_{\rm xc}(n, |\nabla n|, \cdots)$ is some dimensionless functional in the density and its gradients [69] and is often chosen to satisfy various physical constraints. Since many of these constraints cannot be satisfied simultaneously the form of $F_{\rm xc}(n, |\nabla n|, \cdots)$ chosen is the one best suited to the system being studied. The commonest GGA parameterizations used in condensed matter calculations are PW91 [69], [59] and PBE [69].

The GGAs lead to improvements over the LDA in a number of areas including accurate total atomic energies and reduction in the errors in calculating atomisation energies and lattice constants [70]. Bond lengths in molecules are also improved in some cases but are occasionally over- or under-corrected relative to LDA calculations [69]. Despite the extensive advantages in accuracy and applicability of GGAs compared to the LDA, a number of systems, particularly where non-local electron correlations are strong require treatment by other methods. These include LDA+U [45] which is the LDA with an extra, orbital dependent term to deal with the strong Coulomb interaction, and the GW method, where the non-local interactions are included via an electron self-interaction [71].

In the next section we will discuss the extension of DFT to include relativistic effects.

3.2 Relativistic Density Functional Theory

Having discussed standard density functional theory (DFT), which is essentially a non-relativistic theory and the Dirac equation, which is relativistic, it seems sensible to discuss relativistic density functional theory (RDFT). Both DFT and RDFT have a common aim, to reduce the problem of the many-body wavefunction describing a fully interacting system of electrons to a set of single particle equations which can be solved to give some fundamental quantity (in the case of DFT, the electron density) from which we can calculate physical properties. However, as we have seen already, the Dirac equation leads to several subtle considerations which have no direct analog in non-relativistic quantum mechanics.

When discussing the exchange-correlation potential earlier, we introduced a correlation function measuring the probability of finding an electron at $\mathbf{r}_1$ and simul-

taneously finding an electron at $\mathbf{r}_2$. In the non-relativistic case there is no problem with this, but the definition of simultaneity depends on reference frame in a relativistic theory. A further consideration is the electromagnetic interaction between two electrons. In our multi-electron system the electrons are moving charges and there will be electric currents associated with this motion. In a non-relativistic model, these currents interact with each other instantaneously but in a relativistic theory the electromagnetic interaction travels at the speed of light and there will be a time delay (retardation) depending on the separation of the electrons and, of course, the speed of light. In general, there is also a problem with the constraint of non-relativistic DFT that the number of particles remains fixed. In a fully relativistic theory we must allow for the creation of electron positron pairs in keeping with quantum electrodynamics. However, since pair creation is typically a high-energy phenomenon it is reasonably safe to assume that this can be ignored in the context of ground state calculations in condensed matter applications.

A proper treatment of relativistic Density Functional Theory including all the effects of the interactions of electrons and photons theory should begin with the Dirac equation for a system of interacting electrons under the influence of a full electromagnetic four-potential in the framework of quantum electrodynamics. Following through with arguments analogous to Hohenberg and Kohn leads to a description in terms of a total energy functional, $W[J^\mu(\mathbf{r})]$, of the *four current density*, $J^\mu(\mathbf{r})$ [72], [73]:

$$J^\mu(\mathbf{r}) = (J_x(\mathbf{r}), J_y(\mathbf{r}), J_z(\mathbf{r}), cn(\mathbf{r}))\,, \tag{3.37}$$

where the $J_i(\mathbf{r})$ are the components of the electric current, $n(\mathbf{r})$ is the charge density and c is the speed of light. The ground state current density is

$$J^\mu = \langle\Psi|\hat{J}^\mu|\Psi\rangle \tag{3.38}$$

with the components of the current density subject to the conservation of probability density

$$\frac{\partial J^\mu}{\partial x^\mu}, \tag{3.39}$$

where relativistic notation has been introduced: $x^1 = x$, $x^2 = y$, $x^3 = z$, $x^4 = ct$. As in the non-relativistic case, the four-current density which minimises the total energy functional is the ground-state density. There is however an additional consideration. The electromagnetic vector potential, $\mathbf{A}(\mathbf{r})$, is related to the *observable* magnetic field, $\mathbf{B}(\mathbf{r})$ via

$$\mathbf{B}(\mathbf{r}) = -\nabla \times \mathbf{A}(\mathbf{r}), \tag{3.40}$$

and since $\nabla \times \nabla\phi(\mathbf{r}) = 0$ for some scalar function $\phi(\mathbf{r})$, the choice of vector potential is not unique. We can add an arbitrary $\nabla\phi(\mathbf{r})$ and still get the same magnetic field. This is an example of a gauge transformation and the total energy functional is

unique to within this gauge transformation. In other words, unlike non-relativistic DFT where the relationship between the external potential and the ground state density is unique, in the relativistic case we must consider the ground state current density to be related to a unique *set* of potentials related by a gauge transformation.

Having established the connection between the total energy functional and the ground state current density it follows that we should be able to construct a set of single particle equation analogous to the Kohn Sham equations in non-relativistic DFT. If we ignore the pair creation effects we can assume, as before that the number of particles is fixed, leading to the constraint [10]

$$\int \mathrm{d}^3\mathbf{r} n(\mathbf{r}) = N. \tag{3.41}$$

As before, we consider variations in the total energy functional and arrive at the Dirac-Kohn-Sham equations [72], [73], [74]:

$$\left\{ c\boldsymbol{\alpha} \cdot [\hat{\mathbf{p}} - e\mathbf{A}_{\mathrm{eff}}(\mathbf{r})] + \beta mc^2 + V_{\mathrm{eff}}(\mathbf{r}) \right\} \psi_i(\mathbf{r}) = w_i \psi_i(\mathbf{r}), \tag{3.42}$$

where the $\psi_i(\mathbf{r})$ are a set of single particle, 4-component eigenstates and the w_i are Lagrange multipliers serving as eigenvalues. The charge and current densities are given by

$$n(\mathbf{r}) = \sum_{i=1}^{N} \psi_i^\dagger(\mathbf{r})\psi_i(\mathbf{r}), \qquad \mathbf{J}(\mathbf{r}) = c\sum_{i=1}^{N} \psi_i^\dagger(\mathbf{r})\boldsymbol{\alpha}\psi_i(\mathbf{r}) \tag{3.43}$$

where the sums are over the N lowest electron eigenvalues. The effective vector and scalar potentials, $\mathbf{A}_{\mathrm{eff}}(\mathbf{r})$ and $V_{\mathrm{eff}}(\mathbf{r})$, are given by

$$\mathbf{A}_{\mathrm{eff}}(\mathbf{r}) = \mathbf{A}_{\mathrm{ext}} + \int \mathrm{d}^3\mathbf{r}' \frac{\mathbf{J}(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|} + \frac{1}{e}\frac{\delta E_{\mathrm{xc}}[n(\mathbf{r}), \mathbf{J}(\mathbf{r})]}{\delta \mathbf{J}(\mathbf{r})} \tag{3.44}$$

and

$$V_{\mathrm{eff}}(\mathbf{r}) = V_{\mathrm{ext}} + \int \mathrm{d}^3\mathbf{r}' \frac{n(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta E_{\mathrm{xc}}[n(\mathbf{r}), \mathbf{J}(\mathbf{r})]}{\delta n(\mathbf{r})}. \tag{3.45}$$

The relativistic total energy functional is written [10]

$$\begin{aligned} W[n(\mathbf{r}), \mathbf{J}(\mathbf{r})] = \sum_{i=1}^{N} w_i - \frac{1}{2}\int \mathrm{d}^3\mathbf{r} \int \mathrm{d}^3\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{\mathrm{xc}}[n(\mathbf{r}), \mathbf{J}(\mathbf{r})] \\ + \frac{1}{2c^2}\int \mathrm{d}^3\mathbf{r} \int \mathrm{d}^3\mathbf{r}' \frac{\mathbf{J}(\mathbf{r})\mathbf{J}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \\ - \int \mathrm{d}^3 n(\mathbf{r}) \frac{\delta E_{\mathrm{xc}}[n(\mathbf{r}), \mathbf{J}(\mathbf{r})]}{\delta n(\mathbf{r})} + \int \mathrm{d}^3\mathbf{J}(\mathbf{r}) \frac{\delta E_{\mathrm{xc}}[n(\mathbf{r}), \mathbf{J}(\mathbf{r})]}{\delta \mathbf{J}(\mathbf{r})}. \end{aligned} \tag{3.46}$$

The first line is clearly the same as the non-relativistic expression, containing the energy eigenvalues, the Hartree term describing the classical Coulomb interaction and the exchange-correlation energy. The second line is the classical current-current interaction analogous to the Hartree term. The final line arises because the exchange-correlation energy is a functional of both the charge and current densities. The fact that the total energy functional does not depend explicitly on $\mathbf{A}(\mathbf{r})$ means that it does not depend on the choice of gauge for the vector potential. This gauge invariance must hold for all observable quantities hence we would expect it of the total energy.

For magnetic systems, it is possible [72], [74] to derive a density functional theory which takes as its fundamental quantity not the four current density, $[\mathbf{J}(\mathbf{r}), n(\mathbf{r})]$, but the magnetisation and the electron density, $[\mathbf{m}(\mathbf{r}), n(\mathbf{r})]$.

As for the non-relativistic case, the accuracy of RDFT depends on the availability of a good approximation to the exchange correlation energy functional. Again, as in standard DFT, a local density approximation can be made using the exchange correation function for a now relativistic homogeneous electron gas. The exchange part of the exchange-correlation function can be solved analytically [73], [75] and is typically decomposed into contributions from the Coulomb interaction and the transverse component of the electron-photon interaction for the electromagnetic field [73]. The correlation effects are, in general, more difficult to approximate and variety of methods have been employed to generate and implement relativistic exchange correlation functions (see for example [76]). The relativistic local density approximation (RLDA), as in the non-relativistic case, has some justification for applications although tends not to perform so well due to the non-local nature of the correlation part of the exchange correlation potential for the relativistic homogeneous electron gas. Extensions to the RLDA analogous to those of non-relativistic DFT have been made, including a generalised gradient approximation [77]. However, because of the difficulties of relativistic approximations to the exchange correlation functional, it has become common practice in many relativistic electronic structure calculations to use non-relativistic approximations.

In this chapter we have shown that the complicated many-body problem of a large number of interacting electrons can, by the application of density functional theory, be reduced to a set of single particle equations which are solved self-consistently for the electron density.

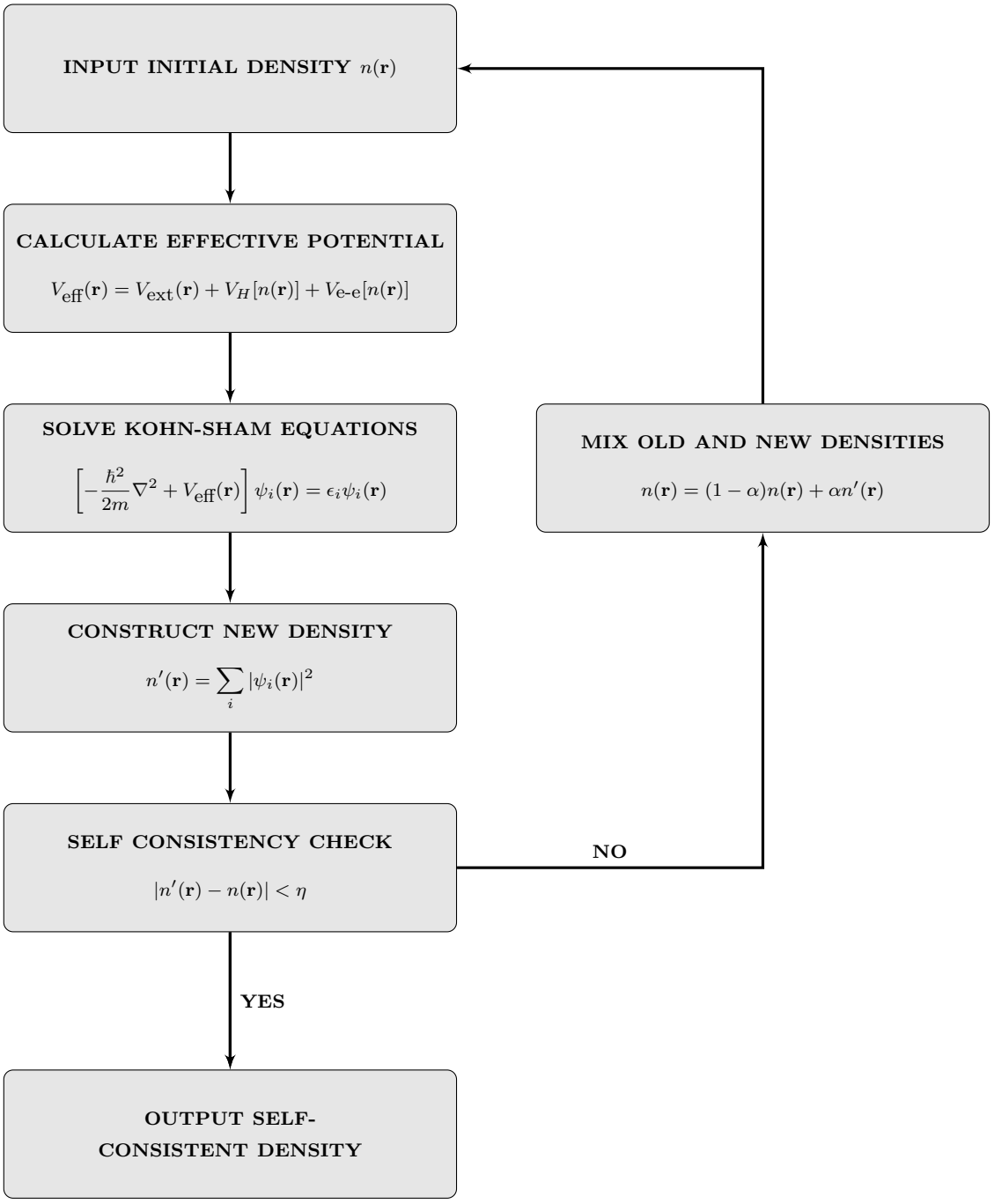

Figure 3-1: A generalised iteration procedure for solving the Kohn-Sham equations self-consistently for the charge density $n(\mathbf{r})$.

Chapter 4

THE GREEN'S FUNCTION EMBEDDING METHOD

This chapter introduces the Green's function embedding method for electronic structure calculations in general terms, beginning by discussing the motivation for embedding by looking at what happens when we break bulk translational symmetry in a small region of space and then comparing some methods for dealing with this. We then discuss in detail the mathematical framework of non-relativistic embedding for general systems and give an example of its implementation for a simple, model system. Following on, we present the framework for relativistic embedding and discuss the similarities and differences between the two, concentrating in particular on basis set expansions of the wavefunction. Again, we provide an example of its implementation for the same simple system. Next comes a brief introduction to self-consistent embedding which is covered in detail in the rest of the thesis, and we conclude with a presentation of the Linearised Augmented Plane Wave basis which has been used in self-consistent embedded surface calculations.

Previously, we have discussed the general single-particle quantum theory of relativistic spin-1/2 particles and the density functional approach to performing practical calculations on the large number of interacting electrons encountered in atoms, molecules and solids. From now on we will focus on determining the electronic structure of crystalline solids within Density Functional Theory.

The ideal crystal is a regular arrangement, or lattice, of atoms extending to infinity in all directions and leads to the idea of infinite translational symmetry, allowing the lattice to be described by a single unit cell containing just a few atoms, considerably reducing the computational expense of calculations. This is possible from Bloch's theorem, which states that the wavefunction must reflect the symmetry of the lattice and enables periodic boundary conditions to be imposed on the wavefunction to mimic the infinite repetition of the unit cell. Many bulk crystal properties can be calculated using this approach.

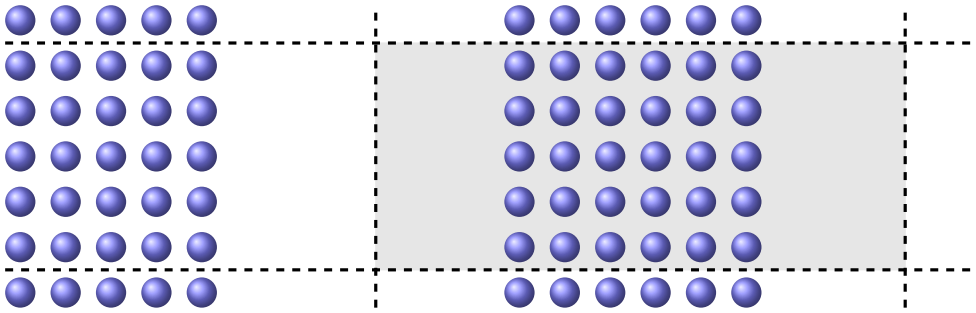

Figure 4-1: The geometry in a typical supercell surface calculation. The shaded area shows a single cell and the dashed lines the cell edges along which periodic boundary conditions are imposed.

As soon as translational symmetry is broken, for example by a surface or impurity, the idea of the unit cell breaks down and periodic boundary conditions are no longer applicable. Fortunately, the change in the potential felt by the electrons due to the surface or impurity differs from the bulk potential only in a small region. This means that the problem can be divided into distinct regions; the bulk, far from the surface or impurity, where we may still use periodic boundary conditions, and the region around the surface or impurity where the potential differs significantly from the bulk, requiring special treatment. Calculations on surfaces can be divided into two classes, supercell and semi-infinite.

4.1 Supercell and Semi-Infinite Calculations

One method of treating systems where translational symmetry is broken in a small region of space is the supercell approach which allows periodic boundary conditions to be retained [78]. To illustrate the basic idea of a supercell we will consider a crystal surface (see figure (4-1). For unreconstructed surfaces, the translational symmetry in the system is broken only in the direction perpendicular to the surface. Parallel to the surface, the translational symmetry is retained and in this direction the calculation may still be performed within a single unit cell. In calculations to determine long-range surface reconstruction the cell may need to be extended parallel to the surface to accommodate the shifts in atomic positions in the surface plane.

The potential usually only differs from the bulk in a few atomic layers below the surface, due to screening, and in a small region of vacuum above it. The number of atomic layers in the cell is chosen so that the potential in the layers at the boundary of the cell below the surface is as close as possible to that of the bulk crystal. A similar condition is imposed on the boundary in the vacuum above the surface. Enough vacuum is chosen so that the potential far from the surface is constant. Similar considerations are necessary when considering a substitutional impurity, vacancy or

defect; along the cell edges the potential must be bulk like.

Supercell methods have been used to study a wide range of systems where translational symmetry is broken, including: surface reconstruction [79], surface adsorption [80], surface defects [81], work functions [82], surface electronic states [83], [16], vacancies, dopants and impurities [84], [85], [86] and electron transport across interfaces [87], [88].

Despite their widespread applicability, supercell methods have certain drawbacks. One disadvantage is that the number of atoms in the cell can be large, leading to calculations which require considerable computing power. Calculation of the (7×7) surface reconstruction of Si requires 1000 atoms per cell [79]. The perturbation of the potential caused by the surface or impurity can be quite extended and the number of atoms needed to recover the bulk-like potential can also be large. For example, an investigation of the spin-orbit split surface state on Au(111) requires 23 atomic layers in the supercell [16]. Aside from the number of atoms in the cell, supercell calculations are somewhat lacking in a true representation of bulk continuum states. The discrete energy levels of atoms coalesce into continuous energy bands in bulk crystals and investigations of the band structure of solids provide a valuable insight into their physical properties. An accurate representation of bulk bands is particularly important for the interpretation of experimental spectroscopies, such as photoemission, transport calculations and understanding localised surface states and resonances. Because of the artificial periodic boundary conditions imposed on supercells, the bulk bands are not a true continuum, but rather a set of discrete states (see fig 4-2). This can make separation of surface resonances from the bulk states difficult [89]. The size of the cell can also affect surface states. Surface states from adjacent interfaces can interact, causing a splitting in energy. When investigating spin-orbit splitting of surface states this can be problematic if the adjacent surface state interaction leads to an energy splitting on the scale of the spin-orbit splitting [90] (see fig 4-2). The size of the cell has to be chosen to minimise the interaction of adjacent surface states.

The second class of methods for treating systems with translational symmetry broken in a small region are those that reproduce the bulk continuum of states. The reason for the more accurate representation of the bulk band structure is that the true infinite translation symmetry is retained in the bulk, rather than the artificial periodic boundary conditions imposed in supercell calculations. In the case of surface calculations, the surface is created by truncating the infinite crystal, resulting in a semi-infinite bulk.

A particularly well developed and widely applied example of a semi-infinite method is the Korringa-Kohn-Rostoker (KKR) formalism based on a multiple scattering approach [91], [92]. The KKR method is primarily a Green's function technique and has been used to perform calculations on a range of systems. It has been

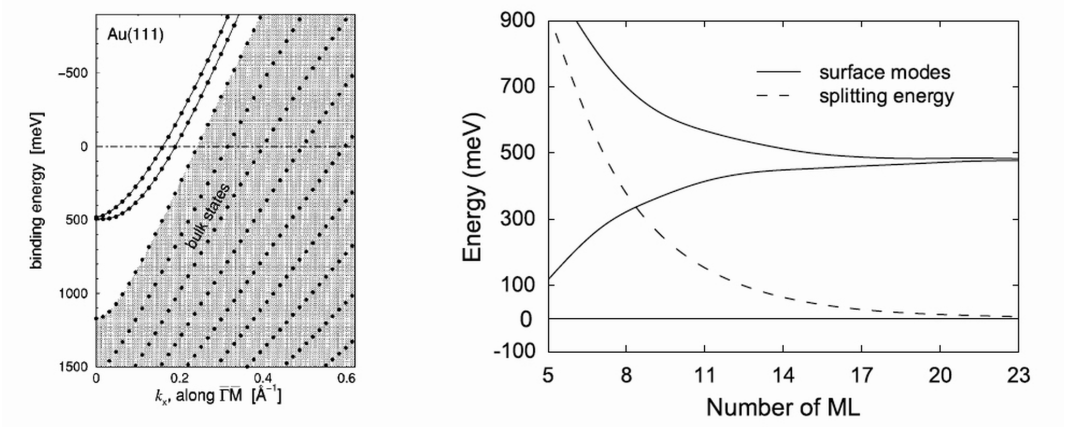

Figure 4-2: Two examples of the disadvantages of supercell methods. The figure on the left shows the Au(111) surface state and the projected bulk band structure (dark grey area) demonstrating the representation of bulk states as discrete rather than a continuum (reproduced from [16]) which, for example, complicates identification of the bulk band edge. The plot on the right is from a different calculation of the dispersion of the Au(111) surface state [90]. It shows the artificial energy splitting due to interaction of surface states from opposite sides of the same cell as a function of the number of Au layers in the cell. The actual spin-orbit splitting of the Au(111) surface state is ≈ 100 meV.

applied to ballistic transport in layered materials [93] and non-collinear magnetic structures [94] and is particularly useful since it can deal with transport in disordered systems [95]. Although it is possible to determine surface relaxations [96] and lattice distortion due to impurities [97], it is not as powerful as the supercell method for general systems, favouring close packed over more open structures or those in which significant lattice reconstructions occur. KKR is very well suited to surface calculations [98], particularly the Layer-KKR (LKKR) extension [99]. The semi-infinite nature of the technique means it is suitable for investigating surface state properties [100], [101], [15]. Although a full potential method exists [102], this can be computationally demanding and a large number of KKR calculations are performed in the muffin-tin approximation; a constant potential between non-overlapping atomic spheres with spherically symmetric atom-centered potentials [45].

A second semi-infinite approach, the Green's Function Embedding Method introduced by Inglesfield [2] is the main topic of this work. In the embedding method, the area in which the potential differs from the bulk, a small region around an impurity or a few atomic layers below a surface, is explicitly treated, while the influence of the bulk crystal is replaced by an energy dependent potential, the *embedding potential*.

A particular advantage of this method is that the embedding potential for a given material need only be calculated once over a given energy range, allowing calculations on clean surfaces and monolayers or different impurities with the same substrate to be implemented efficiently [89]. In addition, for surface calculations, the

number of atomic layers in the embedded region is significantly reduced compared to the equivalent supercell calculation. For example, a calculation of the spin-orbit split surface state dispersion on Au(111) requires 23 Au layers in the supercell [103]. The embedding method has therefore been extensively used in surface calculations to determine surface state dispersions and surface band structures, work functions and charge densities [104], [105], [106], [107], [108], [109], [110]. Because it is a semi-infinite method it is able to distinguish between surface states and resonances [111], [112], [89],and is ideally suited to describing image potential states [113], [114]. A good description of the ground state band structure and Green's function are required for investigation of the optical response of surfaces and the embedding method has been used to provide this as a starting point for certain studies [115], [116].

Accurate bulk and surface band structures are also key in the study of electron transport across interfaces since they describe the available conduction channels [117] and these systems are important in the field of spintronics. Likewise, a proper treatment of conduction channels in the semi-infinite leads connected to a nanodevice is provided by the embedding method. Several implementations of the embedding method have been derived to allow conductivity calculations to be performed accurately in these systems, [46], [118], [119], [120].

In addition to surface and interface calculations, the embedding method has been applied to confined systems [121], [122], adsorbates [123] and surface relaxation and total energies [124]. Other methods are generally preferred for relaxation and total energy calculations over the embedding approach since the surface region is not automatically constrained to be charge neutral, allowing charge transfer between the bulk and surface regions [106].

The rest of this chapter describes the Green's function Embedding Method in detail. First of all the general, non-relativistic version of the method is discussed along with its implementation for embedded self-consistent surface calculations. Various ways of calculating the embedding potential are covered including, in particular the transfer matrix approach which is subsequently incorporated into the relativistic version of embedding and allows the bulk complex band structure to be calculated. In the final part of this chapter the extension of the general non-relativistic embedding method to the Dirac equation is discussed. The remaining chapters of the thesis are devoted to extending the relativistic approach to full-potential surface calculations.

4.2 Non-Relativistic Embedding

We have seen that there are a number of points to be addressed when performing calculations on systems where the infinite translational symmetry of a bulk crystal is broken, specifically that deviation from the bulk potential is usually confined to

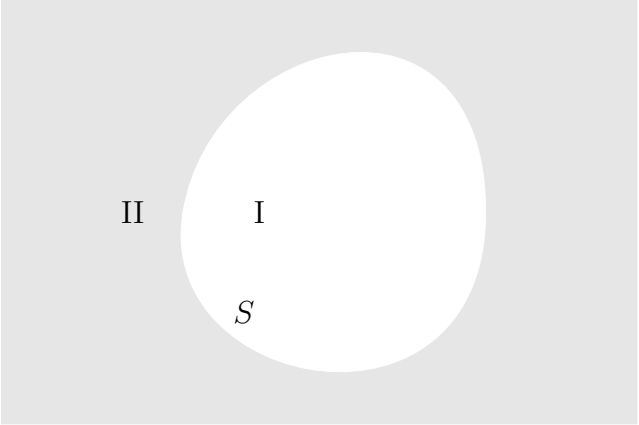

Figure 4-3: In the embedding scheme we consider a system divided into two distinct regions, I and II, separated by a closed surface S.

region close to the impurity or surface responsible for breaking the symmetry. To derive the embedding method we therefore begin with some general system which may be divided into two distinct regions, labelled region I and II, separated by some closed surface S, fig 4-3. By closed, we mean a surface which separates the embedded region, I, completely from the 'bulk' region, II. In the case of an impurity S simply encloses some volume around the impurity and it is easy to see how region I is 'closed'. When modelling a surface (fig 4-4)we typically allow the translational symmetry to extend infinitely in the dimensions parallel to the crystal surface and we define S_L as a plane parallel to the surface which separates off a small number of surface layers from the bulk and S_R a surface which separates off a region of vacuum away from the crystal surface. The embedding surface $S = S_L \cup S_R$ and since S_L and S_R extend to infinity then regions I and II are isolated by S. The division of the problem into regions I and II in this manner applies equally to both non–relativistic and relativistic systems.

The non-relativistic embedding method aims to generate the Green's function for the entire system, I+II, by solving the Schrödinger equation explicitly only in region I. This is achieved by adding an energy dependent surface potential to the Hamiltonian for region I so that solutions of the Schrödinger equation in this region match smoothly on to wavefunctions in region II. To find the form for this surface potential, we begin by defining a trial function,

$$\Psi(\mathbf{r}) = \begin{cases} \phi(\mathbf{r}) & \mathbf{r} \in I \\ \psi(\mathbf{r}) & \mathbf{r} \in II \end{cases} . \tag{4.1}$$

$\phi(\mathbf{r})$, in region I matches in amplitude on to a solution, $\psi(\mathbf{r})$, of the Schrödinger

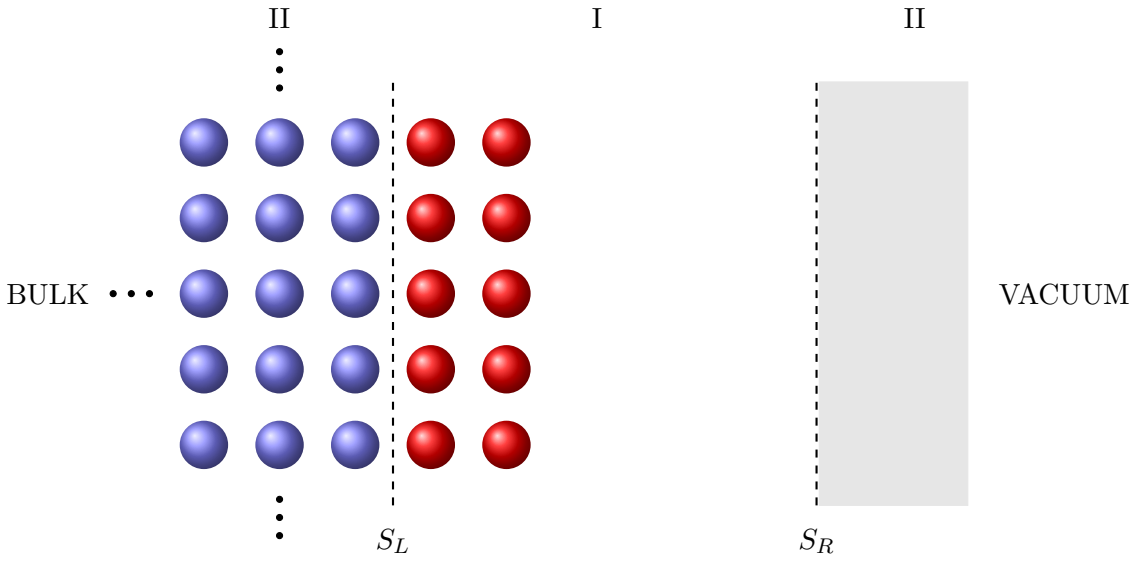

Figure 4-4: For a surface calculation, the closed surface S which separates the embedded region, I, from region II is formed by two surfaces: S_L, dividing the bulk crystal from the surface layers and S_R which divides the vacuum from the embedded region.

equation, at energy ϵ, in region II:

$$\phi(\mathbf{r}_S) = \psi(\mathbf{r}_S) \quad \mathbf{r}_S \text{ on } S, \qquad H_{II}\psi(\mathbf{r}) = \left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})\right]\psi(\mathbf{r}) = \epsilon\psi(\mathbf{r}) \quad \mathbf{r} \in II. \tag{4.2}$$

The trial function is defined so that it is continuous in amplitude over S, but this does not mean that in general the normal derivative will be. Since the wavefunction must be single valued and continuous throughout space, and we will use the embedding method to impose the continuous normal derivative boundary condition.

The energy expectation value of the entire system is

$$E = \frac{\langle\phi|H|\phi\rangle_{I+II}}{\langle\phi|\phi\rangle_{I+II}}. \tag{4.3}$$

but we need to be careful with the kinetic energy terms in the Hamiltonian because of the discontinuity in normal derivative on S. From Green's first identity [34]

$$\int_I \mathrm{d}^3\mathbf{r}\,\phi^*(\mathbf{r})\nabla^2\phi(\mathbf{r}) = -\int_I \mathrm{d}^3\mathbf{r}\,\nabla\phi^*(\mathbf{r})\cdot\nabla\phi(\mathbf{r}) + \int_S \mathrm{d}^2\mathbf{r}_S\cdot\phi^*(\mathbf{r}_S)\nabla\phi(\mathbf{r})$$

$$\int_{II} \mathrm{d}^3\mathbf{r}\,\psi^*(\mathbf{r})\nabla^2\psi(\mathbf{r}) = -\int_{II} \mathrm{d}^3\mathbf{r}\,\nabla\psi^*(\mathbf{r})\cdot\nabla\psi(\mathbf{r}) + \int_S \mathrm{d}^2\mathbf{r}_S\cdot\psi^*(\mathbf{r}_S)\nabla\psi(\mathbf{r}) \tag{4.4}$$

In the surface integrals in (4.4), the area element $\mathrm{d}^2\mathbf{r}_S$ is a vector which is perpen-

dicular to the surface. We define this surface normal to *always* point *out* of region I *into* region II and we will write

$$\int_S \mathrm{d}^2\mathbf{r}_S \cdot \phi^*(\mathbf{r}_S)\nabla\phi(\mathbf{r}) = \int_S \mathrm{d}^2 r_S \phi^*(\mathbf{r}_S)\frac{\partial\phi(\mathbf{r})}{\partial n_S}$$

$$\int_S \mathrm{d}^2\mathbf{r}_S \cdot \psi^*(\mathbf{r}_S)\nabla\psi(\mathbf{r}) = -\int_S \mathrm{d}^2 r_S \psi^*(\mathbf{r}_S)\frac{\partial\psi(\mathbf{r})}{\partial n_S}. \tag{4.5}$$

This means that

$$\int_I \mathrm{d}^3\mathbf{r}\,\phi^*(\mathbf{r})\nabla^2\phi(\mathbf{r}) + \int_{II} \mathrm{d}^3\mathbf{r}\,\psi^*(\mathbf{r})\nabla^2\psi(\mathbf{r}) =$$
$$-\int_I \mathrm{d}^3\mathbf{r}\,\nabla\phi^*(\mathbf{r})\cdot\nabla\phi(\mathbf{r}) - \int_{II} \mathrm{d}^3\mathbf{r}\,\nabla\psi^*(\mathbf{r})\cdot\nabla\psi(\mathbf{r}) + \int_S \mathrm{d}^2 r_S \phi^*(\mathbf{r}_S)\left[\frac{\partial\phi(\mathbf{r})}{\partial n_S} - \frac{\partial\psi(\mathbf{r})}{\partial n_S}\right] \tag{4.6}$$

where we have used the fact that $\phi(\mathbf{r}_S) = \psi(\mathbf{r}_S)$. The final term in (4.6) occurs because we have not specified that the normal derivatives of ϕ and ψ should match on S.

The expectation value of the energy is therefore

$$E = \frac{\int_I \mathrm{d}^3\mathbf{r}\,\phi^*(\mathbf{r})H\phi(\mathbf{r}) + \epsilon\int_{II} \mathrm{d}^3\mathbf{r}\,\psi^*(\mathbf{r})\psi(\mathbf{r}) + \frac{\hbar^2}{2m}\int_S \mathrm{d}^2 r_S\,\phi^*(\mathbf{r}_S)\left[\frac{\partial\phi(\mathbf{r}_S)}{\partial n_S} - \frac{\partial\psi(\mathbf{r}_S)}{\partial n_S}\right]}{\int_I \mathrm{d}^3\mathbf{r}\,\phi^*(\mathbf{r})\phi(\mathbf{r}) + \int_{II} \mathrm{d}^3\mathbf{r}\,\psi^*(\mathbf{r})\psi(\mathbf{r})}, \tag{4.7}$$

where we have used (4.2) for the second term in the numerator.

We now eliminate ψ in (4.7). Firstly we use the Green's function, $G_0(\mathbf{r},\mathbf{r}';\epsilon)$, for region II to express the normal derivative of ψ in terms of ϕ. The Green's function satisfies

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) - \epsilon\right] G_0(\mathbf{r},\mathbf{r}';\epsilon) = -\delta(\mathbf{r}-\mathbf{r}') \qquad \mathbf{r},\mathbf{r}' \in II. \tag{4.8}$$

and the wavefunction in II satisfies

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) - \epsilon\right] \psi(\mathbf{r}) = 0 \qquad \mathbf{r} \in II. \tag{4.9}$$

Multiplying (4.9) by G_0, (4.8) by ψ, subtracting and integrating over region II gives

$$\psi(\mathbf{r}) = -\frac{\hbar^2}{2m}\int_{II} \mathrm{d}^3\mathbf{r}'\left[G_0(\mathbf{r},\mathbf{r}';\epsilon)\nabla^2_{\mathbf{r}'}\psi(\mathbf{r}') - \psi(\mathbf{r})\nabla^2_{\mathbf{r}'}G_0(\mathbf{r},\mathbf{r}';\epsilon)\right] \tag{4.10}$$

where $\nabla^2_{\mathbf{r}'}$ acts on the $\mathbf{r}'$ coordinate. We can write this as a surface integral using

Green's second theorem [34]

$$\psi(\mathbf{r}) = \frac{\hbar^2}{2m}\int_S \mathrm{d}^2 r_s \left[G_0(\mathbf{r},\mathbf{r}_S;\epsilon)\frac{\partial\psi(\mathbf{r}_S)}{\partial n_S} - \psi(\mathbf{r}_S)\frac{\partial G_0(\mathbf{r},\mathbf{r}_S;\epsilon)}{\partial n_S}\right]. \tag{4.11}$$

where the normal derivative acts on the second argument, $\mathbf{r}_S$, in the Green's function. If we choose the Green's function to satisfy the zero normal derivative boundary condition on S:

$$\frac{\partial G_0(\mathbf{r},\mathbf{r}_S;\epsilon)}{\partial n_S} = 0 \tag{4.12}$$

and let $\mathbf{r} \to \mathbf{r}_S$ from within region II, (4.11) we get a relationship between the amplitude and normal derivative of ψ on S:

$$\psi(\mathbf{r}_S) = \frac{\hbar^2}{2m}\int_S \mathrm{d}^2 r'_s\, G_0(\mathbf{r}_S,\mathbf{r}'_S;\epsilon)\frac{\partial\psi(\mathbf{r}'_S)}{\partial n'_S}. \tag{4.13}$$

Since $\psi(\mathbf{r}_S) = \phi(\mathbf{r}_S)$ on $\mathbf{r}_S$, if we invert this equation we can eliminate $\partial\psi(\mathbf{r}_S)/\partial n_S$ in (4.7). We can do this by defining the surface inverse, $G_0^{-1}(\mathbf{r},\mathbf{r}_S;\epsilon)$ of $G_0(\mathbf{r},\mathbf{r}_S;\epsilon)$ by

$$\int_S \mathrm{d}^2 r''_S\, G_0(\mathbf{r}_S,\mathbf{r}''_S;\epsilon)G_0^{-1}(\mathbf{r}''_S,\mathbf{r}'_S;\epsilon) = \delta(\mathbf{r}_S - \mathbf{r}'_S), \tag{4.14}$$

so that

$$\frac{\partial\psi(\mathbf{r}_S)}{\partial n_S} = -\frac{2m}{\hbar^2}\int_S \mathrm{d}^2 r'_S\, G_0^{-1}(\mathbf{r}_S,\mathbf{r}'_S;\epsilon)\phi(\mathbf{r}'_S) \tag{4.15}$$

The next step is to replace the integrals containing $|\psi(\mathbf{r})|^2$. To do this we consider the Schrödinger equation for $\psi(\mathbf{r})$ in II and its complex conjugate:

$$H\psi = \epsilon\psi \tag{4.16}$$

$$H\psi^* = \epsilon\psi^* \tag{4.17}$$

where we assume real energies. Differentiating (4.16) with respect to ϵ we have

$$H\frac{\partial\psi}{\partial\epsilon} = \psi + \epsilon\frac{\partial\psi}{\partial\epsilon} \tag{4.18}$$

We now multiply (4.18) by ψ^* and (4.17) by $\partial\psi/\partial\epsilon$,

$$\psi^* H\frac{\partial\psi}{\partial\epsilon} = |\psi|^2 + \epsilon\psi^*\frac{\partial\psi}{\partial\epsilon} \tag{4.19}$$

$$\frac{\partial\psi}{\partial\epsilon}H\psi^* = \epsilon\psi^*\frac{\partial\psi}{\partial\epsilon}. \tag{4.20}$$

Subtracting, rearranging and substituting in for the Hamiltonian gives

$$|\psi(\mathbf{r})|^2 = \frac{\hbar^2}{2m}\left[\frac{\partial\psi(\mathbf{r}_S)}{\partial\epsilon}\nabla^2\psi^*(\mathbf{r}_S) - \psi^*(\mathbf{r}_S)\nabla^2\frac{\partial\psi(\mathbf{r}_S)}{\partial\epsilon}\right]. \tag{4.21}$$

We now integrate over region II and use Green's second identity (again) to convert the volume integral to a surface integral:

$$\int_{II} \mathrm{d}^3\mathbf{r}\,|\psi(\mathbf{r})|^2 = -\frac{\hbar^2}{2m}\int_S \mathrm{d}^2 r_S\,\left[\frac{\partial\psi(\mathbf{r}_S)}{\partial\epsilon}\frac{\partial\psi^*(\mathbf{r}_S)}{\partial n_S} - \psi^*(\mathbf{r}_S)\frac{\partial}{\partial\epsilon}\frac{\partial\psi(\mathbf{r}_S)}{\partial n_S}\right], \tag{4.22}$$

where the minus sign in front comes from our convention that the normal derivative points out of region I into region II. The first term in the surface integral is zero because, by construction, $\psi(\mathbf{r}_S) = \phi(\mathbf{r}_S)$ and $\phi(\mathbf{r}_S)$ is independent of ϵ. The volume integral is therefore

$$\int_{II} \mathrm{d}^3\mathbf{r}\,|\psi(\mathbf{r})|^2 = \frac{\hbar^2}{2m}\int_S \mathrm{d}^2 r_S\,\psi^*(\mathbf{r}_S)\frac{\partial}{\partial\epsilon}\frac{\partial\psi(\mathbf{r}_S)}{\partial n_S} \tag{4.23}$$

and we can substitute from (4.15) to get

$$\int_{II} \mathrm{d}^3\mathbf{r}\,|\psi(\mathbf{r})|^2 = -\int_S \mathrm{d}^2 r_S \int_S \mathrm{d}^2 r'_S\,\phi^*(\mathbf{r}_S)\frac{\partial G_0^{-1}(\mathbf{r}_S,\mathbf{r}'_S;\epsilon)}{\partial\epsilon}\phi(\mathbf{r}'_S). \tag{4.24}$$

which gives the norm of the trial function in region II in terms of the energy derivative of the surface inverse of the Green's function in II.

Finally, we substitute this into the expression for the energy expectation value and obtain an equation which contains only the trial function in region I, $\phi(\mathbf{r})$, and the surface quantity $G_0^{-1}(\mathbf{r}_S,\mathbf{r}'_S;\epsilon)$,

$$\begin{aligned} E = \Bigg[&\int_I \mathrm{d}^3\mathbf{r}\,\phi^*(\mathbf{r})H\phi(\mathbf{r}) + \frac{\hbar^2}{2m}\int_S \mathrm{d}^2 r_S\,\phi^*(\mathbf{r}_S)\frac{\partial\phi(\mathbf{r}_S)}{\partial n_S} \\ &+ \int_S \mathrm{d}^2 r_S \int_S \mathrm{d}^2 r'_S\,\phi^*(\mathbf{r}_S)G_0^{-1}(\mathbf{r}_S,\mathbf{r}'_S;\epsilon)\phi(\mathbf{r}'_S) \\ &-\epsilon\int_S \mathrm{d}^2 r_S \int_S \mathrm{d}^2 r'_S\,\phi^*(\mathbf{r}_S)\frac{\partial G_0^{-1}(\mathbf{r}_S,\mathbf{r}'S;\epsilon)}{\partial\epsilon}\phi(\mathbf{r}'_S)\Bigg] \Bigg/ \\ &\left[\int_I \mathrm{d}^3\mathbf{r}\,\phi^*(\mathbf{r})\phi(\mathbf{r}) - \int_S \mathrm{d}^2 r_S \int_S \mathrm{d}^2 r'_S\,\phi^*(\mathbf{r}_S)\frac{\partial G_0^{-1}(\mathbf{r}_S,\mathbf{r}'S;\epsilon)}{\partial\epsilon}\phi(\mathbf{r}'_S)\right]. \end{aligned} \tag{4.25}$$

To see how this is a variational method for determining the minimum energy eigenvalue for the whole system, I+II, we consider the change in the expectation

value due to variations $\delta\phi(\mathbf{r})$:

$$\delta E = \left[\int_I \mathrm{d}^3\mathbf{r}\,\delta\phi^*(\mathbf{r})\,[H-E]\,\phi(\mathbf{r}) + \frac{\hbar^2}{2m}\int_S \mathrm{d}^2 r_S\,\delta\phi^*(\mathbf{r}_S)\frac{\partial\phi(\mathbf{r}_S)}{\partial n_S}\right.$$
$$+\int_S \mathrm{d}^2 r_S \int_S \mathrm{d}^2 r'_S\,\delta\phi^*(\mathbf{r}_S)G_0^{-1}(\mathbf{r}_S,\mathbf{r}'_S;\epsilon)\phi(\mathbf{r}'_S)$$
$$\left.+(E-\epsilon)\int_S \mathrm{d}^2 r_S \int_S \mathrm{d}^2 r'_S\,\delta\phi^*(\mathbf{r}_S)\frac{\partial G_0^{-1}(\mathbf{r}_S,\mathbf{r}'_S;\epsilon)}{\partial\epsilon}\phi(\mathbf{r}'_S)\right] /$$

$$\left[\int_I \mathrm{d}^3\mathbf{r}\,\phi^*(\mathbf{r})\phi(\mathbf{r}) - \int_S \mathrm{d}^2 r_S \int_S \mathrm{d}^2 r'_S\,\phi^*(\mathbf{r}_S)\frac{\partial G_0^{-1}(\mathbf{r}_S,\mathbf{r}'_S;\epsilon)}{\partial\epsilon}\phi(\mathbf{r}'_S)\right]. \tag{4.26}$$

If E is stationary with respect to the variations in $\phi^*(\mathbf{r})$ then we have an equation satisfied by the trial function $\phi(\mathbf{r})$

$$E\phi(\mathbf{r}) = \left[-\frac{\hbar^2}{2m}\nabla^2_{\mathbf{r}} + \frac{\hbar^2}{2m}\delta(\mathbf{r}-\mathbf{r}_S)\frac{\partial}{\partial n_S} + V(\mathbf{r})\right]\phi(\mathbf{r})$$
$$+\,\delta(\mathbf{r}-\mathbf{r}_S)\int_S \mathrm{d}^2 r'_S\left[G_0^{-1}(\mathbf{r}_S,\mathbf{r}'_S;\epsilon) + (E-\epsilon)\frac{\partial G_0^{-1}(\mathbf{r}_S,\mathbf{r}'_S;\epsilon)}{\partial\epsilon}\right]\phi(\mathbf{r}'_S). \tag{4.27}$$

This is an effective Schrödinger equation for $\phi(\mathbf{r})$ in which the quantity $G_0^{-1}(\mathbf{r}_S,\mathbf{r}'_S;\epsilon)$ enters as an effective surface potential operating on $\phi(\mathbf{r})$ and is called the *embedding potential*.

From (4.27) we see that,

$$H\phi(\mathbf{r}) = E\phi(\mathbf{r}), \qquad \mathbf{r}\in I \tag{4.28}$$

and

$$\frac{\partial\phi(\mathbf{r}_S)}{\partial n_S} = -\frac{2m}{\hbar^2}\int_S \mathrm{d}^2 r_S\left[G_0^{-1}(\mathbf{r}_S,\mathbf{r}'_S;\epsilon) + (E-\epsilon)\frac{\partial G_0^{-1}(\mathbf{r}_S,\mathbf{r}'_S;\epsilon)}{\partial\epsilon}\right]\phi(\mathbf{r}'_S). \tag{4.29}$$

The first of these shows that the trial function satisfies the Schrödinger equation at energy E in region I. By comparing the second expression with (4.15), we see that $\phi(\mathbf{r})$ also satisfies the correct normal derivative on S when $E=\epsilon$. The term in $\partial G_0^{-1}(\mathbf{r}_S,\mathbf{r}'_S;\epsilon)/\partial\epsilon$ is a first order correction for when ϵ is not equal to the correct energy E.

If we expand the trial function $\phi(\mathbf{r})$ in region I in some set of basis functions

$$\phi(\mathbf{r}) = \sum_n a_n\chi_n(\mathbf{r}), \tag{4.30}$$

then substitute into (4.27) we obtain a matrix equation in the expansion coefficients,

a_n

$$\sum_n \left\{ [H]_{nn'} + \left[G_0^{-1}\right]_{nn'} + (E-\epsilon)\left[\dot{G}_0^{-1}\right]_{nn'} \right\} a_{n'} = E \sum_n [O]_{nn'}\, a_{n'}, \tag{4.31}$$

where the matrix elements are

$$[H]_{nn'} = \int_I \mathrm{d}^3\mathbf{r}\, \chi_n^*(\mathbf{r}) \left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) \right] \chi_{n'}(\mathbf{r}) + \frac{\hbar^2}{2m} \int_S \mathrm{d}^2 r_s \chi_n^*(\mathbf{r}_S) \frac{\partial \chi_{n'}(\mathbf{r}_S)}{\partial n_S} \tag{4.32a}$$

$$\left[G_0^{-1}\right]_{nn'} = \int_S \mathrm{d}^2 r_S \int_S \mathrm{d}^2 r'_S \chi_n^*(\mathbf{r}_S) G_0^{-1}(\mathbf{r}_s, \mathbf{r}'_S; \epsilon) \chi_{n'}(\mathbf{r}'_S) \tag{4.32b}$$

$$\left[\dot{G}_0^{-1}\right]_{nn'} = \int_S \mathrm{d}^2 r_S \int_S \mathrm{d}^2 r'_S \chi_n^*(\mathbf{r}_S) \frac{\partial G_0^{-1}(\mathbf{r}_s, \mathbf{r}'_S; \epsilon)}{\partial \epsilon} \chi_{n'}(\mathbf{r}'_S) \tag{4.32c}$$

$$[O]_{nn'} = \int_I \mathrm{d}^3\mathbf{r} \chi_n^*(\mathbf{r}) \chi_{n'}(\mathbf{r}). \tag{4.32d}$$

In implementing this it is important to ensure that the basis functions (4.30) should be chosen so that they do not satisfy specific boundary conditions on S to ensure sufficient variational freedom. Solving (4.31), the eigenvalues of (4.31) give the energy spectrum of the system and the eigenvectors are the expansion coefficients $\{a_n\}$ of the eigenstates.

4.2.1 The Embedded Green's Function

In the previous section we defined a trial function, $\phi(\mathbf{r})$ in region I which matched in amplitude over the surface S on to a solution of the Schrödinger equation at energy ϵ in region II. The function $\phi(\mathbf{r})$ is the solution of an effective, embedded Schrödinger equation for region I which contains an energy dependent surface potential, $G_0^{-1}(\mathbf{r}_S, \mathbf{r}'_S; \epsilon)$, defined on S. The energy which is stationary with respect to variations in $\phi(\mathbf{r})$ is the correct eigenvalue, E, for I+II. When the embedding potential is evaluated at $\epsilon = E$ in the embedded Schrödinger equation, it ensures that the trial function has the correct normal derivative on S and matches on to the wavefunctions in II at energy E, giving the correct eigenfunctions and eigenvalues of the entire system.

We now consider an embedded system, regions I and II as before, which has a

discrete eigenvalue spectrum. The embedded Schrödinger equation is

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) + \delta(\mathbf{r}-\mathbf{r}_S)\frac{\partial}{\partial n_S}\right]\phi_n(\mathbf{r};E)$$

$$+\delta(\mathbf{r}-\mathbf{r}_S)\int_S \mathrm{d}^2 r'_S G_0^{-1}(\mathbf{r},\mathbf{r}';E)\phi_n(\mathbf{r}'_S;E) = E_n^I(E)\phi_n(\mathbf{r};E) \tag{4.33}$$

where the correction term in the energy derivative of the embedding potential has vanished because we are at the correct energy, E. The eigenfunctions are normalised over region I

$$\int_I \mathrm{d}^3\mathbf{r}\,\phi_n^*(\mathbf{r};E)\phi_{n'}(\mathbf{r};E) = \delta_{nn'} \tag{4.34}$$

and the eigenvalues depend on the energy at which the embedding potential is evaluated; $E_n^I(E)$.

The embedded Green's function for region I, at energy E, is defined by the inhomogeneous embedded Schrödinger equation

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) + \delta(\mathbf{r}-\mathbf{r}_S)\frac{\partial}{\partial n_S} - E\right]G_I(\mathbf{r},\mathbf{r}';E)$$

$$+\delta(\mathbf{r}-\mathbf{r}_S)\int_S \mathrm{d}^2 r''_S G_0^{-1}(\mathbf{r}_S,\mathbf{r}''_S;E)G_I(\mathbf{r}''_S,\mathbf{r}';E) = -\delta(\mathbf{r}-\mathbf{r}') \tag{4.35}$$

and can be written in terms of the eigenfunctions of the homogeneous embedded equation via the spectral representation

$$G_I(\mathbf{r},\mathbf{r}';E) = \sum_n \frac{\phi_n^*(\mathbf{r};E)\phi_n(\mathbf{r}';E)}{E - E_n^I(E)}. \tag{4.36}$$

The Green's function for the entire system, $G_{I+II}(\mathbf{r},\mathbf{r}';E)$ can also be written in the spectral representation

$$G_{I+II}(\mathbf{r},\mathbf{r}';E) = \sum_n \frac{\Psi_n^*(\mathbf{r})\Psi_n(\mathbf{r}')}{E - E_n^{I+II}} \tag{4.37}$$

where E_n^{I+II} are the eigenvalues of the whole system, I+II and the wavefunctions $\Psi_n(\mathbf{r})$ are normalised over I+II

$$\int_{I+II} \mathrm{d}^3\mathbf{r}\,\Psi_n^*(\mathbf{r})\Psi_{n'}(\mathbf{r}) = \delta_{nn'}. \tag{4.38}$$

From (4.36) we can see that the poles of $G_I(\mathbf{r},\mathbf{r}';E)$ are at $E = E_n^I(E)$. Since we have assumed that the embedding potential in (4.33) has been evaluated at the correct energy, it follows that $E_n^I = E_n^{I+II}$ and the poles of G_I and G_{I+II} occur at

the same energies.

For a complex function

$$f(z) = \frac{g(z)}{h(z)} \tag{4.39}$$

with a simple pole at $z = z_0$, the residue at z_0 is given by [48]

$$\mathrm{Res}(z_0) = \frac{g(z_0)}{\left.\dfrac{\partial h(z)}{\partial z}\right|_{z=z_0}} \tag{4.40}$$

so the residues at the poles, E_n^{I+II}, of G_{I+II} are

$$\Psi_n^*(\mathbf{r})\Psi_n(\mathbf{r}'). \tag{4.41}$$

The residues of G_I at $E_n^I(E) = E^{I+II}$ are

$$\frac{\phi_n^*(\mathbf{r})\phi_n(\mathbf{r})}{1 - \dfrac{\partial E_n^I(E)}{\partial E}} = \frac{\phi_n^*(\mathbf{r})\phi_n(\mathbf{r})}{1 - \displaystyle\int_S \mathrm{d}^2 r_S \int_S \mathrm{d}^2 r'_S \phi_n^*(\mathbf{r}_S) \frac{\partial G_0^{-1}(\mathbf{r}_S, \mathbf{r}'_S; E_n)}{\partial E} \phi_n(\mathbf{r}'_S)}. \tag{4.42}$$

where we have simply differentiated the embedded Schrödinger equation with respect to E to get the integral in the denominator. We have encountered this integral, when writing the norm of the trial function in II in terms of the trial function in I (4.24) and the denominator is the factor which corrects for the difference between the normalisation of $\phi_n(\mathbf{r})$ over I+II and the normalisation of $\Psi_n(\mathbf{r})$ over I+II. The residues of G_I are therefore

$$\frac{\phi_n^*(\mathbf{r})\phi_n(\mathbf{r})}{1 - \dfrac{\partial E_n^I(E)}{\partial E}} = \Psi_n^*(\mathbf{r})\Psi_n(\mathbf{r}') \tag{4.43}$$

and these are the residues of G_{I+II}. Thus, the poles and residues of G_I and G_{I+II} are identical, meaning that

$$G_I(\mathbf{r}, \mathbf{r}'; E) = G_{I+II}(\mathbf{r}, \mathbf{r}'; E) \qquad \mathbf{r}, \mathbf{r}' \in I, \tag{4.44}$$

and the embedded Green's function is identical to the Green's function for the entire system evaluated in region I. From now on we will drop the subscript I and write the embedded Green's function as $G(\mathbf{r}, \mathbf{r}'; E)$.

To determine the embedded Green's function we can expand in a set of functions defined in region I, just as we did for the trial function previously,

$$G(\mathbf{r}, \mathbf{r}'; E) = \sum_{nn'} G_{nn'}(E)\chi_n(\mathbf{r})\chi_{n'}^*(\mathbf{r}'). \tag{4.45}$$

Substituting this into the defining equation for the embedded Green's function, mul-

tiplying by $\chi_l^*(\mathbf{r})\chi_{l'}(\mathbf{r}')$ and integrating over $\mathbf{r}$ and $\mathbf{r}'$ through region I we have

$$\sum_{nn'} \left\{ [H]_{ln} + \left[G_0^{-1}(E)\right]_{ln} - E\,[O]_{ln} \right\} G_{nn'}(E)\,[O]_{n'l'} = -\,[O]_{ll'} \tag{4.46}$$

where the matrices $[H]_{ln}$, $\left[G_0^{-1}(E)\right]_{ln}$ and $[O]_{ln}$ are given by (4.32a)-(4.32d). From (4.46) it is clear that

$$\sum_{m} \left\{ [H]_{nm} + \left[G_0^{-1}(E)\right]_{nm} - E\,[O]_{nm} \right\} G_{mn'}(E) = -\delta_{nn'} \tag{4.47}$$

which gives the matrix representation of the embedded Green's function

$$G_{nn'}(E) = \left[EO - G_0^{-1}(E) - H\right]^{-1}_{nn'} \tag{4.48}$$

(compare this with the general definition of the Green's function (2.74)).

At this point it is worth illustrating how the embedding method is put into practice with a simple example.

4.2.2 The Embedded Square Well in One Dimension

As a demonstration of the embedding method we will consider a one-dimensional square well. The potential is

$$V(z) = \begin{cases} -V_0 & \text{for} \quad -\frac{a}{2} \le z \le +\frac{a}{2} \\ \\ 0 & \text{otherwise} \end{cases} \tag{4.49}$$

and the embedded region, region I, is between $-a/2$ and $z = +a/2$ with region II outside the well. In region II, the wavefunctions are free particle solutions to the Schrödinger equation at energy ϵ satisfying suitable boundary conditions at $\pm\infty$:

$$\psi_<(z) = \mathrm{e}^{-\mathrm{i}k_0 z}, \quad z < -\frac{a}{2}, \quad \psi_>(z) = \mathrm{e}^{\mathrm{i}k_0 z}, \quad z \ge +\frac{a}{2}, \quad k_0 = \sqrt{\frac{2m\epsilon}{\hbar^2}}. \tag{4.50}$$

The embedding surface, S is $z = \pm a/2$ and we need to construct an embedding potential to replace region II on this surface. To do this we use the definition of the embedding potential in terms of the normal derivative of the wavefunction in region II (equation (4.15))

$$\frac{\partial\psi(\mathbf{r}_S)}{\partial n_S} = -\frac{2m}{\hbar^2}\int_S \mathrm{d}^2 r'_S\, G_0^{-1}(\mathbf{r}_S, \mathbf{r}'_S; \epsilon)\psi(\mathbf{r}'_S) \tag{4.51}$$

where the surface normal points into region II. Substituting in for the wavefunctions in the one-dimensional system we have

$$G_0^{-1}(\epsilon) = -\frac{\hbar^2}{2m}\mathrm{i}k_0 = -\frac{\hbar^2}{2m}\mathrm{i}\sqrt{\frac{2m\epsilon}{\hbar^2}}, \tag{4.52}$$

the surface integrals reducing to simply evaluating the functions at $\pm a/2$. To demonstrate that the energy expectation value, E, is minimised when the energy, ϵ, at which we evaluate the embedding potential is equal to E, we need to choose a trial function, $\phi(\mathbf{r})$ in region I. For the one-dimensional square well a sensible choice would be

$$\phi(z) = \cos(kz), \qquad k = \sqrt{\frac{2m(E-V)}{\hbar^2}} \tag{4.53}$$

Substituting this into the expression (4.25) for the energy expectation value gives

$$\begin{aligned} E = \Bigg\{ & \left(\frac{\hbar^2}{2m}k^2 + V\right)\int_{-a/2}^{+a/2} \mathrm{d}z\, \cos^2(kz) + \frac{\hbar^2}{2m}\cos\left(\frac{ka}{2}\right)\frac{\mathrm{d}}{\mathrm{d}z}\cos\left(\frac{-ka}{2}\right) \\ & - \frac{1}{2}\cos\left(\frac{-ka}{2}\right)\frac{\mathrm{d}}{\mathrm{d}z}\cos\left(\frac{-ka}{2}\right) + \left[G_0^{-1}(\epsilon) - \epsilon\frac{\partial G_0^{-1}(\epsilon)}{\partial\epsilon}\right] \\ & \times\left[\cos^2\left(\frac{ka}{2}\right) + \cos^2\left(\frac{-ka}{2}\right)\right]\Bigg\} \Bigg/ \\ & \left\{\int_{-a/2}^{+a/2}\mathrm{d}z\, \cos^2(kz) - \frac{\partial G_0^{-1}(\epsilon)}{\partial\epsilon}\left[\cos^2\left(\frac{ka}{2}\right) + \cos^2\left(\frac{-ka}{2}\right)\right]\right\}. \end{aligned} \tag{4.54}$$

Figure 4-5 shows E calculated with a range of values of ϵ for a well of width $a = \pi$ a.u. and depth $V = 5$ a.u. It is clear that the expectation value is minimised for $\epsilon = E = -4.655$a.u.

We now expand the trial function in the well, region I, in a suitable basis. For the square well centred on $z = 0$, solutions are alternately symmetric (+) and antisymmetric (−) about the centre of the well [8] and we choose the basis

$$\phi^{(+)}(z) = \sum_n a_n^{(+)}\cos(k_n^{(+)}z), \quad n \text{ even}, \qquad \phi^{(-)}(z) = \sum_n a_n^{(-)}\sin(k_n^{(-)}z), \quad n \text{ odd}, \tag{4.55}$$

with

$$k_n^{(+)} = \frac{2(n-1)\pi}{d}, \qquad k_n^{(-)} = \frac{2n\pi}{d} \tag{4.56}$$

where the length $d > a$ is chosen so that the basis functions have a range of values on the embedding surface to ensure variational freedom. Both sets of solutions can be treated separately and we will concentrate on the symmetric states, $\phi^{(+)}(z)$.

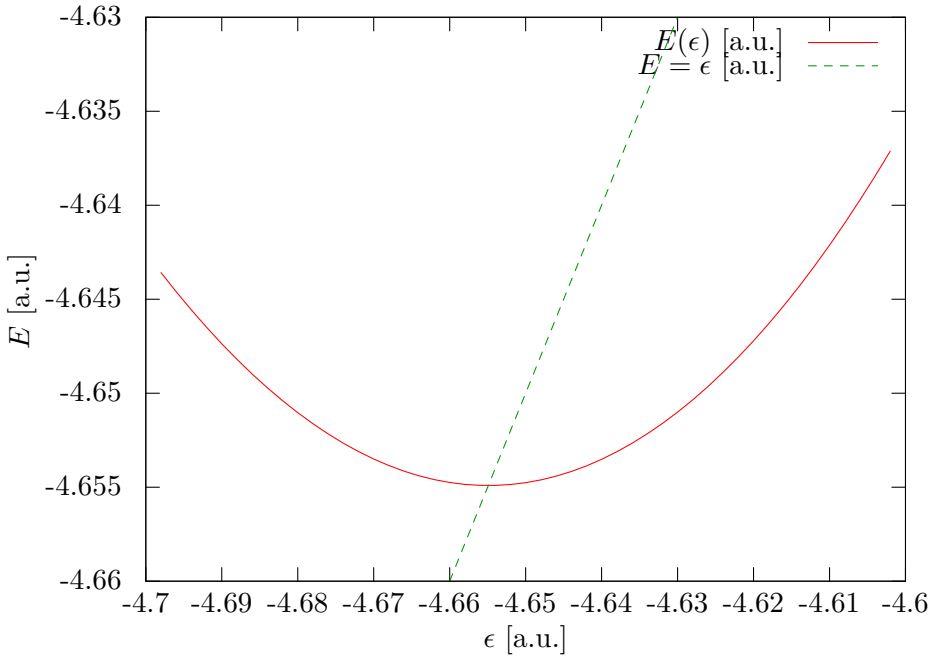

Figure 4-5: Lowest eigenvalue E as a function of energy ϵ for the square well of width $a = \pi$ and depth $V = 5$, showing that the minimum occurs at $E = \epsilon = -4.654913$. All energies are in atomic units (a.u.).

The matrix elements for H, $G_0^{-1}(\epsilon)$, $\dot{G}_0^{-1}(\epsilon)$ and O are calculated from equations (4.32a)-(4.32d).

$$[H]_{nn'} = \left(\frac{\hbar^2}{2m}k_{n'}^2 + V\right)\int_{-a/2}^{+a/2} \mathrm{d}z\, \cos(k_n z)\cos(k_{n'} z) - \frac{\hbar^2}{m}k_{n'}\cos\left(\frac{k_n a}{2}\right)\sin\left(\frac{k_{n'} a}{2}\right) \tag{4.57a}$$

$$\left[G_0^{-1}(\epsilon)\right]_{nn'} = 2G_0^{-1}(\epsilon)\cos\left(\frac{k_n a}{2}\right)\cos\left(\frac{k_{n'} a}{2}\right) \tag{4.57b}$$

$$\left[\dot{G}_0^{-1}(\epsilon)\right]_{nn'} = 2\frac{\partial G_0^{-1}(\epsilon)}{\partial \epsilon}\cos\left(\frac{k_n a}{2}\right)\cos\left(\frac{k_{n'} a}{2}\right) \tag{4.57c}$$

$$[O]_{nn'} = \int_{-a/2}^{+a/2} \mathrm{d}z\, \cos(k_n z)\cos(k_{n'} z) \tag{4.57d}$$

We begin with an initial estimate for ϵ which is somewhere inside the well, evaluate

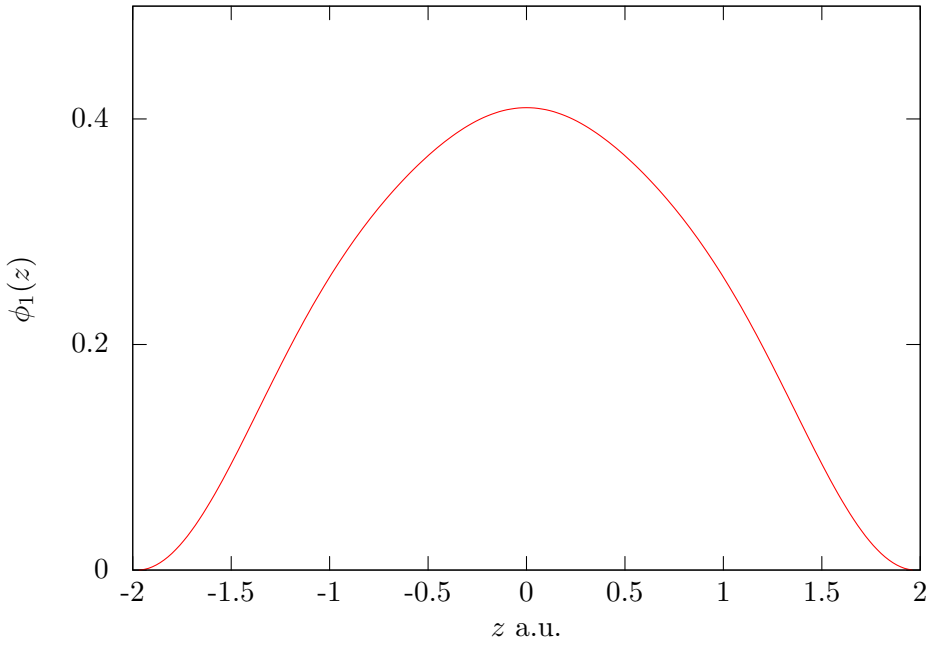

Figure 4-6: The ground state wavefunction, $\phi_1(z)$, for the embedded square well of width $a = \pi$ a.u. and depth $V = 5$ a.u. The wavefunction is expanded in the basis (4.55) with $d = 1.26a$ and $N = 8$, the expansion coefficients given by the eigenvectors of (4.58).

the matrix elements and solve

$$\sum_{n'=1}^{N} \left[H - G_-^{-1}(\epsilon) - \epsilon \dot{G}_0^{-1}(\epsilon) \right]_{nn'} a_{n'} = E_n \sum_{n'=1}^{N} \left[O - \dot{G}_0^{-1}(\epsilon) \right]_{nn'} a_{n'} \tag{4.58}$$

for the eigenvalue, E_n. Table 4.1 shows that as we increase the size of the basis at a fixed ϵ, the eigenvalue approaches the correct value from above. Alternatively, we can keep the size of the basis fixed and calculate E_n self consistently. To do this we start with an initial value of ϵ, solve (4.58) for E_n and then use this E_n as a new input value of ϵ, continuing the cycle until $E_n = \epsilon$. This procedure gives the 'exact' values in table 4.1.

The a_n in (4.58) are the expansion coefficients of the wavefunction in the basis (4.55) at energy E_n. Figure 4-6 shows the lowest energy wavefunction for a well of width $a = \pi$ a.u. and depth $V = -5$ a.u. for a basis of size $N = 8$.

Finally, we find the Green's function for the well, from (4.48)), and use this to calculate the density of states,

$$n(E) = -\frac{1}{\pi} \mathrm{Im} \sum_{n=1}^{N} \sum_{n'=1}^{N} \int_{-a/2}^{+a/2} G_{nn'}(E + \mathrm{i}\eta) \cos(k_n z) \cos(k_{n'} z). \tag{4.59}$$

where a small imaginary part, $\mathrm{i}\eta$ has been added to the energy as discussed in section (2.2). Figure 4-7 shows the density of states for the well using $\eta = 0.01$ a.u. For

N	$\epsilon = V$	$\epsilon = V/2$	exact
2	-4.6283572	-4.7607179	-4.6367002
4	-4.6465276	-4.7298768	-4.6543340
6	-4.6480963	-4.7007652	-4.6548865
8	-4.6499256	-4.6698993	-4.6549120
10	-4.6549129	-4.6549130	-4.6549133

Table 4.1: Values of the lowest eigenvalue, E_1 calculated for different values of N and ϵ for the well with $a = \pi$, $V = 5$ and $d = 1.26a$.

negative energies we have bound states we are in the well and we would expect to see a discrete eigenvalue spectrum. These are clearly visible in the density of states as the sharp peaks at $E = -4.6549$ and $E = -2.0252$, representing the poles in the Green's function. The small imaginary part added to the energy has broadened the delta functions at these poles. For positive energies we have a continuum of states which shows a resonance peak close to $E = 0$. This resonance is a state in the continuum close to the top of the well which will eventually become a bound state as the well deepens.

This simple example has illustrated several things that have been discussed previously. We see that the embedding method correctly reproduces the energy expectation value E when the embedding potential is evaluated at $\epsilon = E$ and at this energy the wavefunction matches in amplitude and derivative on the embedding surface S. The Green's function can be calculated by inverting $EO - H - G_0^{-1}$ and provides an efficient way of determining the density of states. The addition of a small imaginary part to the energy broadens the delta function peaks occurring at the poles of the Green's function for a discrete energy spectrum and allows energy integrals to be performed along the real axis for continuum states.

4.3 Relativistic Embedding

In general terms, the relativistic and non-relativistic embedding methods are the same: we divide our system into two regions, I and II, separated by a closed surface, S, and solve the problem in region I alone, the influence of region II being introduced via an embedding potential defined on S. The differences arise from the fact that

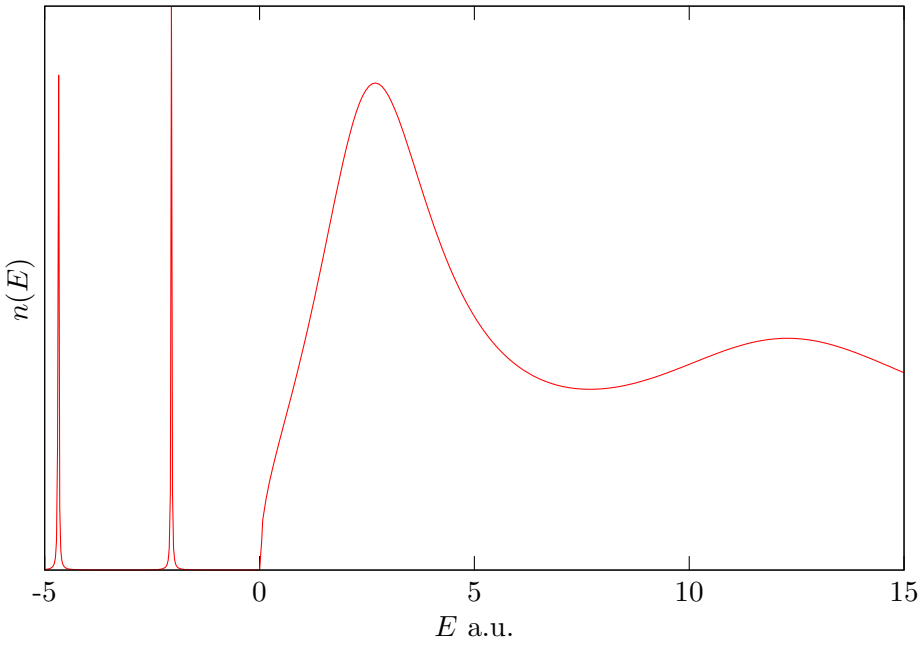

Figure 4-7: Density of states, $n(E)$, for the square well $a = \pi$, $V = 5$, $d = 1.26a$. A small imaginary part, $i\eta$ has been added to the energy with $\eta = 0.01$ a.u.

we solve the Dirac equation in region I, rather than the Schrödinger equation and the wavefunction is a 1×4 spinor and the Green's function is a 4×4 matrix.

Once again, we begin by defining a trial function,

$$\Psi(\mathbf{r}) = \begin{cases} \phi(\mathbf{r}) & \mathbf{r} \in I \\ \psi(\mathbf{r}) & \mathbf{r} \in II \end{cases} \tag{4.60}$$

The large component of the trial function, $\phi(\mathbf{r})$, in region I matches in amplitude on S onto a solution, $\psi(\mathbf{r})$, of the Dirac equation at energy w in region II [3]:

$$\phi_l(\mathbf{r}_S) = \psi_l(\mathbf{r}_S), \qquad \left[-\mathrm{i}c\hbar\boldsymbol{\alpha}\cdot\nabla + \beta mc^2 + V(\mathbf{r}) - w\right]\psi(\mathbf{r}) = 0, \ \mathbf{r} \in II, \tag{4.61}$$

where we have neglected the vector potential, $\mathbf{A}(\mathbf{r})$ for clarity and the energy, w, contains the rest energy, mc^2. In the non-relativistic case, we chose the amplitude of the trial function to be continuous across S and allowed the derivative to be discontinuous. The non-relativistic wavefunction must be continuous in amplitude and first derivative throughout the whole system because of the second order nature of the Schrödinger equation. The Dirac equation is first order and the requirement is that the amplitude of both large and small components are continuous over the whole system. Here we have chosen the amplitude of the large component to be continuous but have no such requirement for the small component. In the non-relativistic case, the embedding potential, when evaluated at the correct energy, ensured that the normal derivative of the trial function was continuous throughout the system. We

therefore expect the relativistic embedding potential to perform a similar role for the amplitude of the small component.

The energy expectation value for I+II is

$$W = \frac{\langle\phi|H|\phi\rangle_{I+II}}{\langle\phi|\phi\rangle_{I+II}} \tag{4.62}$$

Care needs to be taken with the kinetic energy term, $-\mathrm{i}c\hbar\boldsymbol{\alpha}\cdot\nabla$, because the small components of ϕ and ψ can be discontinuous on S. Writing the trial function as

$$\chi(\mathbf{r}) = \phi(\mathbf{r})\Theta_I + \psi(\mathbf{r})\Theta_{II}, \tag{4.63}$$

where

$$\Theta_I = \begin{cases} 1 & \mathbf{r} \in I \\ 0 & \mathbf{r} \in II \end{cases}, \qquad \Theta_{II} = \begin{cases} 1 & \mathbf{r} \in II \\ 0 & \mathbf{r} \in I \end{cases}, \tag{4.64}$$

then

$$\begin{aligned}\nabla\chi(\mathbf{r}) &= \nabla\phi(\mathbf{r})\,\Theta_I + \phi(\mathbf{r})\nabla\Theta_I + \nabla\psi(\mathbf{r})\,\Theta_{II} + \psi(\mathbf{r})\nabla\Theta_{II} \\ &= \nabla\phi(\mathbf{r})\,\Theta_I - \phi(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}_S) + \nabla\psi(\mathbf{r})\,\Theta_{II} + \psi(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}_S).\end{aligned} \tag{4.65}$$

The kinetic energy term in (4.62) involves

$$\begin{aligned}\int_{I+II}\mathrm{d}^3\mathbf{r}\,\chi^\dagger(\mathbf{r})\boldsymbol{\alpha}\cdot\nabla\chi(\mathbf{r}) &= \int_I \mathrm{d}^3\mathbf{r}\,\phi^\dagger(\mathbf{r})\boldsymbol{\alpha}\cdot\nabla\phi(\mathbf{r}) + \int_{II}\mathrm{d}^3\mathbf{r}\,\psi^\dagger(\mathbf{r})\boldsymbol{\alpha}\cdot\nabla\psi(\mathbf{r}) \\ &\quad + \int_S \mathrm{d}^2\mathbf{r}_S\cdot\chi^\dagger(\mathbf{r}_S)\boldsymbol{\alpha}\left[\psi(\mathbf{r}_S)-\phi(\mathbf{r}_S)\right].\end{aligned} \tag{4.66}$$

If we write the 4×4 matrix $\boldsymbol{\alpha}$ in the surface integral in terms of the 2×2 Pauli matrices, $\boldsymbol{\sigma}$, and use the matching condition $\phi(\mathbf{r}_S) = \psi(\mathbf{r}_S)$ we get

$$\int_S \mathrm{d}^2\mathbf{r}_S\cdot\chi^\dagger(\mathbf{r}_S)\boldsymbol{\alpha}\left[\psi(\mathbf{r}_S)-\phi(\mathbf{r}_S)\right] = -\int_S \mathrm{d}^2\mathbf{r}_S\cdot\phi_l^\dagger(\mathbf{r}_S)\boldsymbol{\sigma}\left[\phi_s(\mathbf{r}_S)-\psi_s(\mathbf{r}_S)\right], \tag{4.67}$$

where subscripts l and s on wavefunctions denote large and small components respectively. The expectation value is therefore

$$W = \frac{\langle\phi|H|\phi\rangle_I + \langle\psi|H|\psi\rangle_{II} + \mathrm{i}c\hbar\displaystyle\int_S \mathrm{d}^2\mathbf{r}_S\cdot\phi_l^\dagger(\mathbf{r}_S)\boldsymbol{\sigma}\left[\phi_s(\mathbf{r}_S)-\psi_s(\mathbf{r}_S)\right]}{\langle\phi|\phi\rangle_I + \langle\psi|\psi\rangle_{II}} \tag{4.68}$$

and we now eliminate $\psi(\mathbf{r}_S)$.

The trial function in region II satisfies the Dirac equation

$$-\mathrm{i}c\hbar\boldsymbol{\alpha}\cdot\nabla\psi(\mathbf{r})+\left[\beta mc^2+V(\mathbf{r})-w\right]\psi(\mathbf{r})=0,\ \mathbf{r}\in II, \tag{4.69}$$

and if we differentiate this with respect to w we arrive at an equation satisfied by the energy derivative of $\psi(\mathbf{r})$

$$-\mathrm{i}c\hbar\boldsymbol{\alpha}\cdot\nabla\frac{\partial\psi(\mathbf{r})}{\partial w}+\left[\beta mc^2+V(\mathbf{r})-w\right]\frac{\partial\psi(\mathbf{r})}{\partial w}=\psi(\mathbf{r}),\ \mathbf{r}\in II, \tag{4.70}$$

Hermitian conjugating (4.69), multiplying from the right by $\dot{\psi}(\mathbf{r})=\partial\psi/\partial w$, multiplying (4.70) from the left by $\psi^\dagger(\mathbf{r})$ and subtracting gives

$$\psi^\dagger(\mathbf{r})\psi(\mathbf{r})=-\mathrm{i}c\hbar\left[\nabla\psi^\dagger(\mathbf{r})\cdot\boldsymbol{\alpha}\dot{\psi}(\mathbf{r})+\psi^\dagger(\mathbf{r})\boldsymbol{\alpha}\cdot\nabla\dot{\psi}(\mathbf{r})\right] \tag{4.71}$$

where we have assumed that the energy, w, and potential $V(\mathbf{r})$ are real. Integrating over region II we can convert the volume integral into a surface integral

$$\begin{aligned}\langle\psi|\psi\rangle_{II}&=-\mathrm{i}c\hbar\int_{II}\mathrm{d}^3\mathbf{r}\left[\nabla\psi^\dagger(\mathbf{r})\cdot\boldsymbol{\alpha}\dot{\psi}(\mathbf{r})+\psi^\dagger(\mathbf{r})\boldsymbol{\alpha}\cdot\nabla\dot{\psi}(\mathbf{r})\right]\\&=-\mathrm{i}c\hbar\int_{II}\mathrm{d}^3\mathbf{r}\nabla\cdot\left[\psi^\dagger(\mathbf{r})\boldsymbol{\alpha}\dot{\psi}(\mathbf{r})\right]\\&=-\mathrm{i}c\hbar\int_S\mathrm{d}^2\mathbf{r}_S\cdot\psi^\dagger(\mathbf{r}_S)\boldsymbol{\alpha}\dot{\psi}(\mathbf{r}_S).\end{aligned} \tag{4.72}$$

The $\boldsymbol{\alpha}$ can be expanded in Pauli $\boldsymbol{\sigma}$ to give

$$-\mathrm{i}c\hbar\int_S\mathrm{d}^2\mathbf{r}_S\cdot\psi^\dagger(\mathbf{r}_S)\boldsymbol{\alpha}\dot{\psi}(\mathbf{r}_S)=-\mathrm{i}c\hbar\int_S\mathrm{d}^2\mathbf{r}_S\cdot\left[\psi_l^\dagger(\mathbf{r}_S)\boldsymbol{\sigma}\dot{\psi}_s(\mathbf{r}_S)+\psi_s^\dagger(\mathbf{r}_S)\boldsymbol{\sigma}\dot{\psi}_l(\mathbf{r}_S)\right]. \tag{4.73}$$

The second term in this integral is zero because the large component of $\psi(\mathbf{r})$ matches onto the large component of $\phi(\mathbf{r})$ on S and $\phi(\mathbf{r})$ is independent of w, therefore $\dot{\psi}_l(\mathbf{r}_S)=0$. We now have an expression for the normalisation of $\psi(\mathbf{r})$ over region II in terms of the surface values of ϕ_l and $\dot{\psi}_s$:

$$\langle\psi|\psi\rangle_{II}=\mathrm{i}c\hbar\int_S\mathrm{d}^2r_S\,\phi_l^\dagger(\mathbf{r}_S)\sigma_S\frac{\partial\psi_s(\mathbf{r}_S)}{\partial w}, \tag{4.74}$$

where the $-$ sign has been removed, with the surface normal is now taken to point *into* region II. $\sigma_S=\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}_S$ is the projection of $\boldsymbol{\sigma}$ along the normal to S. The corresponding non-relativistic expression is (4.23)

$$\langle\psi|\psi\rangle_{II}=\frac{\hbar^2}{2m}\int_S\mathrm{d}^2r_S\,\phi^*(\mathbf{r}_S)\frac{\partial}{\partial\epsilon}\frac{\partial\psi(\mathbf{r}_S)}{\partial n_S}, \tag{4.75}$$

and we can see that the normal derivative of ψ is replaced by the small component in the relativistic case and complex conjugation is replaced by Hermitian conjugation due to the spinor nature of the wavefunctions.

The next step is to find a way of writing the surface value of the small component of ψ in terms of ϕ. To do this we use the Green's function

$$\left[-\mathrm{i}c\hbar\boldsymbol{\alpha}\cdot\nabla_{\mathbf{r}}+\beta mc^2+V(\mathbf{r})-w\right]G(\mathbf{r},\mathbf{r}';w)=-\delta(\mathbf{r}-\mathbf{r}')\quad \mathbf{r},\mathbf{r}'\in II. \tag{4.76}$$

Multiplying the Hermitian conjugate of this equation from the left by ψ, multiplying (4.69) by $G^\dagger(\mathbf{r},\mathbf{r}';w)$ from the right and subtracting gives

$$\delta(\mathbf{r}-\mathbf{r}')\psi(\mathbf{r})=-\mathrm{i}c\hbar\left[\nabla_{\mathbf{r}}G^\dagger(\mathbf{r},\mathbf{r}';w)\cdot\boldsymbol{\alpha}\psi(\mathbf{r})+G^\dagger(\mathbf{r},\mathbf{r}';w)\boldsymbol{\alpha}\cdot\nabla_{\mathbf{r}}\psi(\mathbf{r})\right]$$

$$=-\mathrm{i}c\hbar\nabla_{\mathbf{r}}\cdot\left[G^\dagger(\mathbf{r},\mathbf{r}';w)\boldsymbol{\alpha}\psi(\mathbf{r})\right], \tag{4.77}$$

where we have assumed real energies. Integrating over region II we can turn the volume integral into a surface integral

$$\psi(\mathbf{r})=-\mathrm{i}c\hbar\int_S \mathrm{d}^2\mathbf{r}_S\cdot G^\dagger(\mathbf{r}_S,\mathbf{r};w)\boldsymbol{\alpha}\psi(\mathbf{r}_S),\quad \mathbf{r}\in II. \tag{4.78}$$

If we write the Green's function in the spectral representation

$$G(\mathbf{r},\mathbf{r}';w)=\sum_n\frac{\chi_n(\mathbf{r})\chi_n^\dagger(\mathbf{r}')}{w-w_n} \tag{4.79}$$

then, at real energies,

$$G^\dagger(\mathbf{r},\mathbf{r}';w)=G(\mathbf{r}',\mathbf{r};w). \tag{4.80}$$

which is the *reciprocity relation*. Using this we can see that the Green's function relates the values of the wavefunction in region II to its amplitude on S:

$$\psi(\mathbf{r})=-\mathrm{i}c\hbar\int_S \mathrm{d}^2\mathbf{r}_S\cdot G(\mathbf{r},\mathbf{r}_S;w)\boldsymbol{\alpha}\psi(\mathbf{r}_S). \tag{4.81}$$

Now writing the Green's function in its large and small components

$$G(\mathbf{r},\mathbf{r}';w)=\begin{pmatrix} G_{ll}(\mathbf{r},\mathbf{r}';w) & G_{ls}(\mathbf{r},\mathbf{r}';w)\\ G_{sl}(\mathbf{r},\mathbf{r}';w) & G_{ss}(\mathbf{r},\mathbf{r}';w)\end{pmatrix} \tag{4.82}$$

where each entry is a 2×2 matrix, and letting $\mathbf{r}\to\mathbf{r}_S$ from within region II we have a relationship between the amplitudes of the large and small components of ψ on S

$$\psi_s(\mathbf{r}_S)=\mathrm{i}c\hbar\int_S \mathrm{d}^2r'_S\,\Gamma(\mathbf{r}_S,\mathbf{r}'_S;w)\sigma_S\psi_l(\mathbf{r}'_S), \tag{4.83}$$

where we have again used the fact that the surface normal points into region II to eliminate the minus sign. In (4.83), we have introduced the 2×2 quantity $\Gamma(\mathbf{r}_S, \mathbf{r}'_S; w)$ defined by

$$\Gamma(\mathbf{r}_S, \mathbf{r}'_S; w) = G_{ss}(\mathbf{r}_S, \mathbf{r}'_S; w) + \mathrm{i}c\hbar \int_S \mathrm{d}^2\mathbf{r}''_S \cdot G_{sl}(\mathbf{r}_S, \mathbf{r}''_S)\boldsymbol{\sigma}\Gamma(\mathbf{r}''_S, \mathbf{r}'_S; w). \tag{4.84}$$

(4.83) is the relativistic analog of the non-relativistic expression (4.13)

$$\psi(\mathbf{r}_S) = \frac{\hbar^2}{2m} \int_S \mathrm{d}^2 r'_s \, G_0(\mathbf{r}_S, \mathbf{r}'_S; \epsilon) \frac{\partial \psi(\mathbf{r}'_S)}{\partial n'_S}. \tag{4.85}$$

from which we can identify $\Gamma(\mathbf{r}_S, \mathbf{r}'_S; w)$ as the embedding potential and we also see again that the normal derivative is replaced by the small component of the wavefunction.

The expectation value W can now be expressed only in terms of the trial function in region I:

$$W = \frac{\langle\phi|H|\phi\rangle_I + \mathrm{i}c\hbar \int_S \mathrm{d}^2 r_S \, \phi_l^\dagger \sigma_S \phi_s + c^2\hbar^2 \int_S \mathrm{d}^2 r_S \int_S \mathrm{d}^2 r'_S \phi_l^\dagger \sigma_S \left[\Gamma - w\dot{\Gamma}\right] \sigma'_S \phi_l}{\langle\phi|\phi\rangle_I - c^2\hbar^2 \int_S \mathrm{d}^2 r_S \int_S \mathrm{d}^2 r'_S \phi_l^\dagger \sigma_S \Gamma \sigma'_S \phi_l} \tag{4.86}$$

which is identical in form to the non-relativistic expression but the small component of the trial function replaces the normal derivative. Just as we did before we can derive an effective Dirac equation by considering changes in W which are minimised with respect to variations in $\delta\phi^\dagger$ and we see that

$$H\phi(\mathbf{r}) = W\phi(\mathbf{r}), \qquad \mathbf{r} \in I \tag{4.87}$$

and

$$\phi_s(\mathbf{r}_S) = \mathrm{i}c\hbar \int_S \mathrm{d}^2\mathbf{r}'_S \cdot \left[\Gamma(\mathbf{r}_S, \mathbf{r}'_S; w) + (W - w)\frac{\partial\Gamma(\mathbf{r}_S, \mathbf{r}'_S; w)}{\partial w}\right] \boldsymbol{\sigma}\phi_l(\mathbf{r}'_S) \tag{4.88}$$

In other words, the trial function satisfies the Dirac equation at energy W in region I and $\phi(\mathbf{r})$ has the appropriate small component on S, matching correctly to the wavefunction $\psi(\mathbf{r})$ in region II. As before, the energy derivative of Γ provides a first order correction when the energy w is not the correct energy W.

4.3.1 The Non-Relativistic Limit of the Embedding Potential

At this point it may be instructive to look at the non-relativistic limit of the embedding potential. We can do this by writing the small component of the wavefunction,

$\psi_s(\mathbf{r})$ in terms of the large component as

$$\psi_s(\mathbf{r}) = \frac{-\mathrm{i}c\hbar}{W - V(\mathbf{r}) + mc^2}\boldsymbol{\sigma}\cdot\nabla\psi_l(\mathbf{r}) \tag{4.89}$$

and substitute this into the equation relating large and small components via the embedding potential:

$$\frac{-\mathrm{i}c\hbar}{W - V(\mathbf{r}) + mc^2}\boldsymbol{\sigma}\cdot\nabla\psi_l(\mathbf{r}_S) = \mathrm{i}c\hbar\int_S \mathrm{d}^2\mathbf{r}'_S\cdot\Gamma(\mathbf{r}_S,\mathbf{r}'_S;W)\boldsymbol{\sigma}\psi_l(\mathbf{r}'_S). \tag{4.90}$$

Multiplying both sides by $-\mathrm{i}c\hbar$ and rearranging we have

$$\boldsymbol{\sigma}\cdot\nabla\psi_l(\mathbf{r}_S) = -\frac{W - V(\mathbf{r}) + mc^2}{c^2\hbar^2}\int_S \mathrm{d}^2\mathbf{r}'_S\cdot c^2\hbar^2\Gamma(\mathbf{r}_S,\mathbf{r}'_S;W)\boldsymbol{\sigma}\psi_l(\mathbf{r}'_S). \tag{4.91}$$

Letting $c\to\infty$ and separating out the rest energy, $W\to E+mc^2$, we see that

$$\boldsymbol{\sigma}\cdot\nabla\psi_l(\mathbf{r}_S) = -\frac{2m}{\hbar^2}\int_S \mathrm{d}^2\mathbf{r}'_S\cdot c^2\hbar^2\Gamma(\mathbf{r}_S,\mathbf{r}'_S;W)\boldsymbol{\sigma}\psi_l(\mathbf{r}'_S). \tag{4.92}$$

We can multiply both sides by the projection of $\boldsymbol{\sigma}$ normal to S, $\boldsymbol{\sigma}\cdot\hat{\mathbf{n}}_S = \sigma_S$ and use the properties of the $\boldsymbol{\sigma}$ matrices to get

$$\frac{\partial\psi_l(\mathbf{r}_S)}{\partial n_S} + \mathrm{i}\boldsymbol{\sigma}\cdot[\hat{\mathbf{n}}_S\times\nabla\psi_l(\mathbf{r}_S)] = -\frac{2m}{\hbar^2}\int_S \mathrm{d}^2 r'_S c^2\hbar^2\sigma_S\Gamma(\mathbf{r}_S,\mathbf{r}'_S;W)\sigma'_S\psi_l(\mathbf{r}'_S). \tag{4.93}$$

The second term on the left hand side comes from a property of defining the spin direction for Dirac wavefunctions. In non-relativistic quantum mechanics, the spin operator is defined in terms of the Pauli matrices

$$\mathbf{S} = \frac{\hbar}{2}\boldsymbol{\sigma} \tag{4.94}$$

and the two-component Pauli spinors, ϕ_{m_s} are eigenfunctions of $\boldsymbol{\sigma}$. Because of the Heisenberg uncertainty principle, we can only precisely measure one component of the spin and the choice of σ_z to be diagonal means that this is the z-component; z is the spin quantisation axis. In relativistic quantum mechanics, the $\mathbf{S}$ operator can be generalised to 4-component spinors

$$\mathbf{S}_4 = \frac{\hbar}{2}\begin{pmatrix}\boldsymbol{\sigma} & 0\\ 0 & \boldsymbol{\sigma}\end{pmatrix}. \tag{4.95}$$

In the rest frame of the electron $\mathbf{S}_4$ distinguishes between spin up and spin down eigenstates, but if the electron is moving pure spin eigenstates only exist when the spin is quantised along the direction of motion [10]. The extra term in (4.93) arises because of coupling between the spin and momentum in relativistic quantum mechanics.

If we ignore the extra term for now, and compare the non-relativistic definition of the embedding potential, we see that, since the large component of the wavefunction is the non-relativistic wavefunction in the $c \to \infty$ limit, (4.93) is identical to the non-relativistic expression if

$$c^2\hbar^2\sigma_S\Gamma(\mathbf{r}_S, \mathbf{r}'_S; W)\sigma'_S \to G_0^{-1}(\mathbf{r}_S, \mathbf{r}'_S; E). \tag{4.96}$$

4.3.2 The Relativistic Embedded Green's Function

Following the presentation of non-relativistic embedding we will now consider the relativistic embedded Green's function. When $w = W$ we can write the embedded Dirac equation, as

$$H\phi(\mathbf{r}) + \int_I \mathrm{d}^3\mathbf{r}'\Delta(\mathbf{r}, \mathbf{r}'; W)\phi(\mathbf{r}') - W\phi(\mathbf{r}) = 0, \tag{4.97}$$

where

$$\Delta(\mathbf{r}, \mathbf{r}'; W) = \delta(\mathbf{r} - \mathbf{r}_S)\delta(\mathbf{r}' - \mathbf{r}'_S) \begin{bmatrix} c^2\hbar^2\sigma_S\Gamma(\mathbf{r}_S, \mathbf{r}'_S; W)\sigma'_S & \mathrm{i}c\hbar\sigma_S\delta(\mathbf{r}_S - \mathbf{r}'_S) \\ 0 & 0 \end{bmatrix}. \tag{4.98}$$

The corresponding Green's function satisfies

$$HG(\mathbf{r}, \mathbf{r}'; W) + \int_I \mathrm{d}^3\mathbf{r}'\Delta(\mathbf{r}, \mathbf{r}'; W)G(\mathbf{r}, \mathbf{r}'; W) - WG(\mathbf{r}, \mathbf{r}'; W) = -\delta(\mathbf{r} - \mathbf{r}') \tag{4.99}$$

and we can show that this coincides with $G_{I+II}(\mathbf{r}, \mathbf{r}'; W)$, the Green's function for the whole system, when $\mathbf{r}, \mathbf{r} \in I$. As before, for simplicity, we consider a finite system with a real and discrete eigenvalue spectrum and write $G_{I+II}(\mathbf{r}, \mathbf{r}'; W)$ in the spectral representation

$$G_{I+II}(\mathbf{r}, \mathbf{r}'; W) = \sum_n \frac{\Psi_n(\mathbf{r})\Psi_n^\dagger(\mathbf{r}')}{W - W_n^{I+II}} \tag{4.100}$$

where the eigenvalues W_n^{I+II} correspond to the eigenstates $\Psi_n(\mathbf{r})$ normalised over I+II. The embedded Green's function in (4.99) can be expanded in a similar fashion

$$G(\mathbf{r}, \mathbf{r}'; W) = \sum_n \frac{\phi_n(\mathbf{r}; W)\phi_n^\dagger(\mathbf{r}'; W)}{W - W_n^I(W)} \tag{4.101}$$

where, $\phi_n(\mathbf{r}; W)$, the eigenfunctions of (4.97) are normalised over region I. As in the non-relativistic case, the poles of G and G_{I+II} are identical because when the embedding potential is evaluated at energy W, the eigenvalues $W_n^I(W)$ are the correct eigenvalues of the entire system, W_n^{I+II}. Next we investigate the residues, as before,

and for $G(\mathbf{r},\mathbf{r}';W)$ these are given by

$$\frac{\phi_n(\mathbf{r};W)\phi_n^\dagger(\mathbf{r}';W)}{1-\left.\dfrac{\partial W_n(W)}{\partial W}\right|_{W_n}} = \frac{\phi_n(\mathbf{r};W)\phi_n^\dagger(\mathbf{r}';W)}{1-c^2\hbar^2\displaystyle\int_S \mathrm{d}^2 r_S \int_S \mathrm{d}^2 r'_S \phi_l^\dagger(\mathbf{r}_S)\sigma_S\Gamma(\mathbf{r}_S,\mathbf{r}'_S;W)\sigma'_S\phi_l(\mathbf{r}'_S)}. \tag{4.102}$$

From (4.83)

$$\langle\psi|\psi\rangle_{II} = -c^2\hbar^2\int_S \mathrm{d}^2 r_S \int_S \mathrm{d}^2 r'_S \phi_l^\dagger(\mathbf{r}_S)\sigma_S\Gamma(\mathbf{r}_S,\mathbf{r}'_S;W)\sigma'_S\phi_l(\mathbf{r}'_S) \tag{4.103}$$

so that the term in the denominator of (4.102) ensures that the trial function is correctly normalised over I+II and

$$\frac{\phi_n(\mathbf{r};W)\phi_n^\dagger(\mathbf{r}';W)}{1-\left.\dfrac{\partial W_n(W)}{\partial W}\right|_{W_n}} = \Psi(\mathbf{r})\Psi^\dagger(\mathbf{r}'). \tag{4.104}$$

The poles of $G(\mathbf{r},\mathbf{r}';W)$ and $G_{I+II}(\mathbf{r},\mathbf{r}';W)$ are therefore identical and have the same weight, so that the two Green's functions coincide when $\mathbf{r},\mathbf{r}' \in I$, just as in the non-relativistic case.

4.3.3 Basis Set Expansions

In the previous section on non-relativistic embedding we obtained the solution of the embedded problem by expanding the trial function in region I in a set of basis functions, deriving matrix equations in the expansion coefficients for the energy eigenvalues, eigenfunctions and the Green's function. We can proceed in exactly the same way in the relativistic case, only now we have a 4-component spinor wavefunction and a 4×4 Green's function. We also have an eigenvalue spectrum which is unbounded below, due to the negative square root in $\pm\sqrt{c^2p^2+m^2c^4}$, and extra care must be taken when converging to the lowest electron eigenvalue in the positive part of the spectrum.

The trial function in region I is expanded in a basis of separate large, $\chi_n^l(\mathbf{r})$, and small component, $\chi_n^s(\mathbf{r})$ 2-spinors

$$\phi(\mathbf{r}) = \sum_{n=1}^{N_l} a_n^l \begin{pmatrix} \chi_n^l(\mathbf{r}) \\ 0 \end{pmatrix} + \sum_{n=1}^{N_s} a_n^s \begin{pmatrix} 0 \\ \chi_n^s(\mathbf{r}) \end{pmatrix} = \begin{bmatrix} \boldsymbol{\chi}^l(\mathbf{r}) & 0 \\ 0 & \boldsymbol{\chi}^s(\mathbf{r}) \end{bmatrix} \begin{bmatrix} \mathbf{a}^l \\ \mathbf{a}^s \end{bmatrix}. \tag{4.105}$$

In general, we don't need the same number of large and small component functions and in the expansion above we have N_l large component spinors and N_s small component spinors and there are N_l+N_s expansion coefficients. The final term in the expansion contains a $4\times(N_l+N_s)$ matrix of basis functions where $\boldsymbol{\chi}^l(\mathbf{r})$ denotes $\chi_1^l(\mathbf{r}),\chi_2^l(\mathbf{r}),\cdots,\chi_{N_l}^l(\mathbf{r})$ etc. and the column vector contains the expansion

coefficients with $\mathbf{a}^l = a^l_1, a^l_2, \cdots, a^l_{N_l}$ etc.

As before, we substitute the expansion (4.105) into the expression for the expectation value W (4.86) and consider W which are stationary with respect to variations in the expansion coefficients and obtain the matrix equation

$$\begin{bmatrix} \mathbf{H}_{ll} & \mathbf{H}_{ls} \\ \mathbf{H}_{sl} & \mathbf{H}_{ss} \end{bmatrix} \begin{bmatrix} \mathbf{a}^l \\ \mathbf{a}^s \end{bmatrix} = W \begin{bmatrix} \mathbf{O}_{ll} & 0 \\ 0 & \mathbf{O}_{ss} \end{bmatrix} \begin{bmatrix} \mathbf{a}^l \\ \mathbf{a}^s \end{bmatrix}. \tag{4.106}$$

The matrix elements are given by

$$\begin{aligned}[H_{ll}]_{nn'} = &\int_I \mathrm{d}^3\mathbf{r}\, \chi^{l\,\dagger}_n(\mathbf{r}) \left[V(\mathbf{r}) + mc^2\right] \chi^l_{n'}(\mathbf{r}) \\ &+ c^2\hbar^2 \int_S \mathrm{d}^2 r_S \int_S \mathrm{d}^2 r'_S \chi^{l\,\dagger}_n(\mathbf{r}_S) \left[\Gamma(\mathbf{r}_S, \mathbf{r}'_S; w) - w\dot{\Gamma}(\mathbf{r}_S, \mathbf{r}'_S; w)\right] \chi^l_{n'}(\mathbf{r}'_S)\end{aligned} \tag{4.107a}$$

$$[H_{ls}]_{nn'} = -\mathrm{i}c\hbar \int_I \mathrm{d}^3\mathbf{r}\, \chi^{l\,\dagger}_n(\mathbf{r}) \boldsymbol{\sigma} \cdot \nabla \chi^s_{n'}(\mathbf{r}) + \mathrm{i}c\hbar \int_S \mathrm{d}^2 r_S\, \chi^{l\,\dagger}_n(\mathbf{r}_S) \sigma_S \chi^s_{n'}(\mathbf{r}_S) \tag{4.107b}$$

$$[H_{sl}]_{nn'} = -\mathrm{i}c\hbar \int_I \mathrm{d}^3\mathbf{r}\, \chi^{s\,\dagger}_n(\mathbf{r}) \boldsymbol{\sigma} \cdot \nabla \chi^l_{n'}(\mathbf{r}) \tag{4.107c}$$

$$[H_{ss}]_{nn'} = \int_I \mathrm{d}^3\mathbf{r}\, \chi^{s\,\dagger}_n(\mathbf{r}) \left[V(\mathbf{r}) - mc^2\right] \chi^s_{n'}(\mathbf{r}) \tag{4.107d}$$

$$\begin{aligned}[O_{ll}]_{nn'} = &\int_I \mathrm{d}^3\mathbf{r}\, \chi^{l\,\dagger}_n(\mathbf{r}) \chi^l_{n'}(\mathbf{r}) \\ &- c^2\hbar^2 \int_S \mathrm{d}^2 r_S \int_S \mathrm{d}^2 r'_S \chi^{l\,\dagger}_n(\mathbf{r}_S) \dot{\Gamma}(\mathbf{r}_S, \mathbf{r}'_S; w) \chi^l_{n'}(\mathbf{r}'_S)\end{aligned} \tag{4.107e}$$

$$[O_{ss}]_{nn'} = \int_I \mathrm{d}^3\mathbf{r}\, \chi^{s\,\dagger}_n(\mathbf{r}) \chi^s_{n'}(\mathbf{r}). \tag{4.107f}$$

By analogy with the non-relativistic case, a similar expansion can be made for the Green's function

$$G(\mathbf{r}, \mathbf{r}'; W) = \begin{bmatrix} \boldsymbol{\chi}^l(\mathbf{r}) & 0 \\ 0 & \boldsymbol{\chi}^s(\mathbf{r}) \end{bmatrix} \mathbf{G}(W) \begin{bmatrix} \boldsymbol{\chi}^l(\mathbf{r}) & 0 \\ 0 & \boldsymbol{\chi}^s(\mathbf{r}) \end{bmatrix}^\dagger, \tag{4.108}$$

which leads to the following expression for the matrix of coefficients, $\mathbf{G}(W)$,

$$[\mathbf{G}(W)]_{nn'} = \begin{bmatrix} WO_{ll} - H_{ll} & -H_{ls} \\ -H_{sl} & WO_{ss} - H_{ss} \end{bmatrix}^{-1}_{nn'} \tag{4.109}$$

with the matrix elements given by (4.107a)-(4.107f) with $\dot{\Gamma} = 0$. In this basis, the density of states is calculated from the coefficients $\mathbf{G}(W)$ and the overlap matrix $\mathbf{O}$

$$n(W) = -\frac{1}{\pi}\mathrm{Im}\,\mathrm{Tr}\,\mathbf{G}(W + \mathrm{i}\eta)\mathbf{O}, \tag{4.110}$$

where, as before, a small imaginary part, $\mathrm{i}\eta$, is added to the energy to broaden the delta functions at bound states.

So far, the relativistic embedding method is very similar to the non-relativistic case, apart from the 4-component nature of wavefunctions and Green's functions. There is, however, an extra consideration that must be made regarding the negative part of the eigenvalue spectrum of the Dirac equation. The embedding method is a variational technique and converges from above to the lowest energy eigenvalue of the system in the non-relativistic case [121]. The Schrödinger equation has negative energy eigenvalues corresponding to bound states, but we have seen that the Dirac equation has a spectrum of negative energy states associated with the negative square root in $W = \pm\sqrt{c^2p^2 + m^2c^4}$ corresponding to positron solutions. There is a forbidden energy range extending from $+mc^2$ to $-mc^2$ for free particles but bound states may exist within this. Bound electron states may therefore have energies $0 < W < mc^2$ and positron-like bound states may have energies $0 > W > -mc^2$. The existence of the negative spectrum means that variational methods of finding eigenvalues of the Dirac equation experience a problem known as *variational collapse* or *finite basis set disease*, [125], [126].

The problem of variational collapse occurs when the Dirac Hamiltonian operator, $\hat{H}$, is represented as a matrix, $\mathbf{H}$ in some finite basis set [126], [127]. Despite what its name suggests, variational collapse does not mean that any variational scheme for solving the Dirac equation results in eigenvalues running away to $-\infty$, but rather the existence of spurious states which lie close to or are degenerate with true eigenvalues [128], [129], [130]. This results in convergence to eigenvalues which are lower than the desired energies [131] and can prevent systematic convergence as the basis set size increases, making accurate numerical calculations hard to achieve.

To see the mechanism which causes variational collapse, we write the Dirac equation in large and small components,

$$\begin{pmatrix} (\hat{V} + mc^2)\mathbb{I}_2 & c\hat{\boldsymbol{\pi}} \\ c\hat{\boldsymbol{\pi}} & (\hat{V} - mc^2)\mathbb{I}_2 \end{pmatrix} \begin{pmatrix} \psi_l \\ \psi_s \end{pmatrix} = W \begin{pmatrix} \psi_l \\ \psi_s \end{pmatrix} \tag{4.111}$$

where $\mathbb{I}_2$ is the 2×2 identity matrix and $\boldsymbol{\pi} = \boldsymbol{\sigma} \cdot \hat{\mathbf{p}}$. The operator $\hat{\boldsymbol{\pi}}$ satisfies the

identity

$$\begin{aligned}\hat{\boldsymbol{\pi}}^2 &= \boldsymbol{\sigma}\cdot\hat{\mathbf{p}}\boldsymbol{\sigma}\cdot\hat{\mathbf{p}} \\ &= \hat{\mathbf{p}}^2 + \mathrm{i}\boldsymbol{\sigma}\cdot(\hat{\mathbf{p}}\times\hat{\mathbf{p}}) \\ &= \hat{\mathbf{p}}^2. \end{aligned} \tag{4.112}$$

If the wavefunction is expanded in a finite basis of N_l large components and N_s small components as in the previous section

$$\psi = \begin{bmatrix} \boldsymbol{\psi}_l \\ \boldsymbol{\psi}_s \end{bmatrix} = \begin{bmatrix} \boldsymbol{\chi}_l & 0 \\ 0 & \boldsymbol{\chi}_s \end{bmatrix} \begin{bmatrix} \mathbf{a}_l \\ \mathbf{a}_s \end{bmatrix} \tag{4.113}$$

and in this basis the Dirac Hamiltonian becomes

$$\mathbf{H} = \begin{bmatrix} \mathbf{V}_{ll} + mc^2\mathbb{I}_l & c\boldsymbol{\pi}_{ls} \\ c\boldsymbol{\pi}_{sl} & \mathbf{V}_{ss} - mc^2\mathbb{I}_s \end{bmatrix} \tag{4.114}$$

where the 'hats' have disappeared because we now have matrix representations of operators and $\mathbb{I}_l$ is the $N_l \times N_l$ identity matrix etc. The solutions ψ of the Dirac equation are now the eigenvectors of

$$\mathbf{H}\begin{bmatrix} \mathbf{a}_l \\ \mathbf{a}_s \end{bmatrix} = W\begin{bmatrix} \mathbf{a}_l \\ \mathbf{a}_s \end{bmatrix} \tag{4.115}$$

for which the energy eigenvalues, W, are stationary with respect to variations in the expansion coefficients $\mathbf{a}$. If we separate out the rest energy and put $E = W - mc^2$ then the large and small components satisfy the matrix equations

$$E\mathbf{a}_l = c\boldsymbol{\pi}_{ls}\mathbf{a}_s + \mathbf{V}_{ll}\mathbf{a}_l \tag{4.116}$$

$$\mathbf{a}_s = c\left[(2mc^2 + E)\mathbb{I}_s - \mathbf{V}_{ss}\right]^{-1}\boldsymbol{\pi}_{sl}\mathbf{a}_l. \tag{4.117}$$

Substituting for $\mathbf{a}_s$ gives an equation for the large components

$$E\mathbf{a}_l = \frac{1}{2m}\boldsymbol{\pi}_{ls}\left[\mathbb{I}_s + \frac{E\mathbb{I}_s - \mathbf{V}_{ss}}{2mc^2}\right]^{-1}\boldsymbol{\pi}_{sl}\mathbf{a}_l + \mathbf{V}_{ll}\mathbf{a}_l. \tag{4.118}$$

This is equivalent to equation (2.35) of section 2.1.4 and if $E\mathbb{I}_s - \mathbf{V}_{ss}$ is small compared to $2mc^2$ we can approximate (4.118) by

$$E\mathbf{a}_l = \frac{1}{2m}\boldsymbol{\pi}_{ls}\boldsymbol{\pi}_{sl}\mathbf{a}_l + \mathbf{V}_{ll}\mathbf{a}_l + \mathcal{O}\left(\frac{1}{c^2}\right). \tag{4.119}$$

which is the non-relativistic limit of the Dirac equation in the basis of large and small

component functions. If the basis of small components is infinite, $N_s \to \infty$, then any small component eigenstate of the operator $\hat{H}$ can be expanded in the basis $\boldsymbol{\chi}_s$ but in a finite basis this is not the case and the basis is incomplete. As a result, the finite set of $\boldsymbol{\chi}_s$ will be a subset of the infinite, complete set of small component basis functions. There will be a set of small component eigenstates of $\hat{H}$, the complement space, which cannot be described by the $\boldsymbol{\chi}_s$ in the finite basis [132], [41]. ‡ If we write an eigenstate in this complement space as $\boldsymbol{\chi}_C$ then [127]

$$\boldsymbol{\pi}_{ls}\boldsymbol{\pi}_{sl} = \mathbf{p}_{ll}^2 - \boldsymbol{\pi}_{lC}\boldsymbol{\pi}_{Cl}, \tag{4.120}$$

the final term going to zero as the small component basis set becomes complete. This means that the representation of the kinetic energy term in the Dirac Hamiltonian does not approach the appropriate non-relativistic limit in a finite basis of small components. In fact, the term $\boldsymbol{\pi}_{lC}\boldsymbol{\pi}_{Cl}$ is always positive [133], [127] and so the calculated kinetic energy is always less than the true kinetic energy. This kinetic energy deficit is the main cause of variational collapse [128], [126].

Various methods are used to prevent variational collapse (see [126]) but the most appropriate in terms of its applicability to the embedding method is *kinetic balance*, since it still allows convergence to eigenvalues from above as the basis set is increased [127] [133]. A kinetically balanced basis is one in which the number of large and small component basis functions is equal, $N_l = N_s = N$, and [127]

$$\boldsymbol{\chi}_s = \boldsymbol{\sigma} \cdot \mathbf{p}\boldsymbol{\chi}_l \tag{4.121}$$

. To see how kinetic balance works we will investigate the effect on the convergence of the electron eigenvalues by increasing the size of the basis set, first by adding an extra small component and then restoring the balance by adding an extra large component. The addition of an extra small component will add an extra eigenvalue to the negative energy spectrum which will raise the positive energy eigenvalues [127]. This is a consequence of Macdonald's theorem†, which states that an extra row and column added to a Hermitian matrix causes the old eigenvalues to interlace, or lie between, the new eigenvalues [135]. Restoring the balance by adding an extra large component basis function introduces a new positive energy eigenvalue which lowers the existing eigenvalues (see figure 4-8). If the downward shift caused by expansion of the large component basis is larger than the upward shift due to the extra small component then we will still converge on the lowest eigenvalue from above as required.

‡An example of this are the $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ unit vectors of 3-dimensional Cartesian space which form a complete set because any vector can be expressed in the basis $(\mathbf{i}, \mathbf{j}, \mathbf{k})$. The subspace $(\mathbf{i}, \mathbf{j})$ is not complete since an arbitrary 3-dimensional vector cannot be expressed only in this basis and the subspace spanned by $\mathbf{k}$ is the complement of $(\mathbf{i}, \mathbf{j})$.

†This itself is related to the Cauchy interlace theorem for real, symmetric matrices [134].

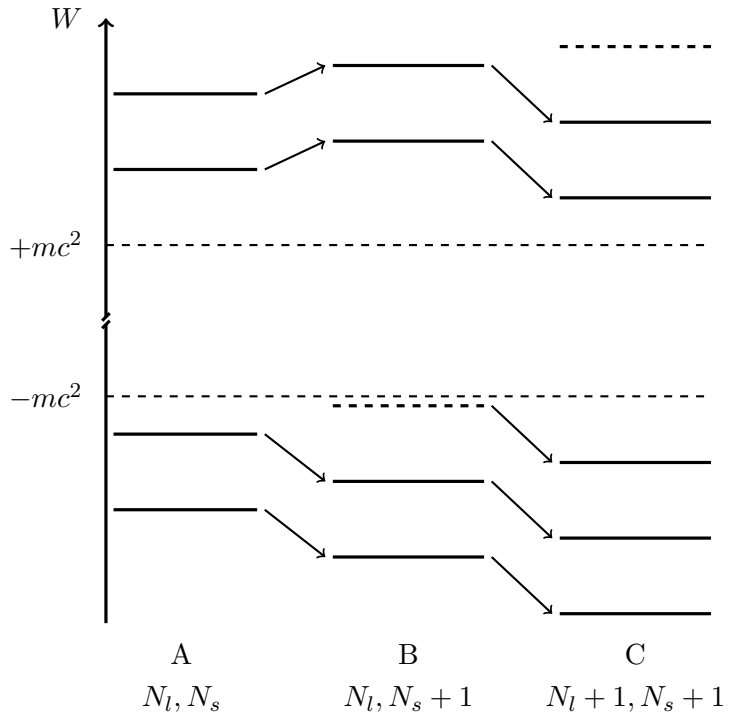

Figure 4-8: Interlacing of eigenvalues due to stepwise increase of the number of large and small component basis functions. (A): Original eigenvalues with $N_l = N_s$. (B): Increasing the small component basis by one, $N_s \to N_s + 1$, introduces a new eigenvalue in the negative spectrum (dashed line) and forces the electron eigenvalues upwards. (C): Restoring the balance by adding an extra large component function, $N_l \to = N_l + 1 = N_s + 1$, causes a downward shift of the electron energies due to the appearance a new eigenvalue.

To qualitatively investigate the relative shifts we consider some general Dirac Hamiltonian $\hat{H} = c\boldsymbol{\alpha}\cdot\hat{\mathbf{p}} + \beta mc^2 + V$ with a scalar potential V but no vector potential. The general matrix eigenvalue problem is

$$\begin{bmatrix} \mathbf{H}_{ll} & \mathbf{H}_{ls} \\ \mathbf{H}_{sl} & \mathbf{H}_{ss} \end{bmatrix} \begin{bmatrix} \mathbf{a}^l \\ \mathbf{a}^s \end{bmatrix} = W \begin{bmatrix} \mathbf{a}^l \\ \mathbf{a}^s \end{bmatrix}. \tag{4.122}$$

Basis functions are

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\psi}^l \\ \boldsymbol{\psi}^s \end{bmatrix}, \qquad \boldsymbol{\psi}^l = \sum_{n=1}^{N_l} a_n^l \chi_n^l, \quad \boldsymbol{\psi}^s = \sum_{n=1}^{N_s} a_n^s \chi_n^s \tag{4.123}$$

and we begin with a kinetically balanced basis where $N_l = N_s = N$ and $\chi_n^s = \boldsymbol{\sigma}\cdot\mathbf{p}\chi_n^l$. The inclusion of an extra small component basis function adds an extra row and column to the Hamiltonian

$$\mathbf{H} \to \mathbf{H} + \mathbf{h}, \qquad \mathbf{h} = \left[\begin{array}{c|c} 0 & h_{ls} \\ \hline h_{sl} & h_{ss} \end{array}\right] = \left[\begin{array}{c|c} 0 & c\boldsymbol{\sigma}\cdot\mathbf{p} \\ \hline c\boldsymbol{\sigma}\cdot\mathbf{p} & V - mc^2 \end{array}\right] \tag{4.124}$$

and the first order change in the m^{th} eigenvalue, from perturbation theory, is $W_m^{(1)} = \langle m|\mathbf{h}|m\rangle$ [8] where $|m\rangle$ is an eigenvector in the original basis without the extra small component.

$$W_m^{(1)} = \langle m|\mathbf{h}|m\rangle = \left[\begin{array}{c|c} \boldsymbol{\Psi}_m^{l\,\dagger} & 0 \end{array}\right] \left[\begin{array}{c|c} 0 & h_{ls} \\ \hline h_{sl} & h_{ss} \end{array}\right] \left[\begin{array}{c} \boldsymbol{\Psi}_m^l \\ \hline 0 \end{array}\right] = 0 \tag{4.125}$$

and there is no change to first order. The second order change is

$$W_m^{(2)} = \frac{|\langle N+1|\mathbf{h}|m\rangle|^2}{W_m^{(0)} - W_{N+1}^{(0)}}, \quad n+1 \neq m \tag{4.126}$$

where $W_m^{(0)}$ is the expectation value of H in the old basis and $W_{n+1}^{(0)}$ is the expectation value of $H + h$ in the new basis.

$$\langle N+1|\mathbf{h}|m\rangle = \left[\begin{array}{c|c} 0 & \chi_{N+1}^{s\,\dagger} \end{array}\right] \left[\begin{array}{c|c} 0 & h_{ls} \\ \hline h_{sl} & h_{ss} \end{array}\right] \left[\begin{array}{c} \boldsymbol{\psi}_m^l \\ \hline \boldsymbol{\psi}_m^s \end{array}\right]$$

$$= c\int \mathrm{d}^3\mathbf{r}\, \chi_{N+1}^{s\,\dagger}\, \boldsymbol{\sigma}\cdot\mathbf{p}\boldsymbol{\psi}_m^l + \int \mathrm{d}^3\mathbf{r}\, \chi_{N+1}^{s\,\dagger}\left[V - mc^2\right]\boldsymbol{\psi}_m^s. \tag{4.127}$$

For a kinetically balanced basis, $\chi_n^s = \boldsymbol{\sigma}\cdot\mathbf{p}\chi_n^l$ and the first integral is

$$\sum_m^{N_l} a_m^l \int \mathrm{d}^3\mathbf{r}\, \chi_{n+1}^{s\,\dagger}\boldsymbol{\sigma}\cdot\mathbf{p}\chi_m^l = \sum_m^{N_l} a_m^l \int \mathrm{d}^3\mathbf{r}\, \chi_{n+1}^{s\,\dagger}\chi_m^s = 0, \quad N+1 \neq m \tag{4.128}$$

if the basis functions are orthogonal. The term in mc^2 also vanishes for orthogonal basis functions and we are left with

$$\langle N+1|\mathbf{h}|m\rangle = \int \mathrm{d}^3\mathbf{r}\, \chi_{N+1}^{s\,\dagger}\, V\, \boldsymbol{\psi}_m^s. \tag{4.129}$$

From (4.117), the small component basis functions are solutions of

$$\boldsymbol{\psi}^s = c\left[(2mc^2 + E) - V\right]^{-1}\boldsymbol{\sigma}\cdot\mathbf{p}\boldsymbol{\psi}^l \tag{4.130}$$

which is of order $1/c$, hence the integral in (4.129) is also order $1/c$ and the second order change in the energy (4.126) is of order $1/c^2$.

For an extra large component, the change in the Hamiltonian is

$$\mathbf{h} = \left[\begin{array}{c|c} h_{ll} & h_{ls} \\ \hline h_{sl} & 0 \end{array}\right] = \left[\begin{array}{c|c} V + mc^2 & c\boldsymbol{\sigma}\cdot\mathbf{p} \\ \hline c\boldsymbol{\sigma}\cdot\mathbf{p} & 0 \end{array}\right]. \tag{4.131}$$

The first order change is again zero

$$W_m^{(1)} = \left[\begin{array}{c|c} 0 & \boldsymbol{\Psi}_m^{l\,\dagger} \end{array}\right] \left[\begin{array}{c|c} h_{ll} & h_{ls} \\ \hline h_{sl} & 0 \end{array}\right] \left[\begin{array}{c} 0 \\ \hline \boldsymbol{\Psi}_m^{l} \end{array}\right] = 0 \tag{4.132}$$

and the second order change reduces to the integral

$$\int \mathrm{d}^3\mathbf{r}\, \chi_{N+1}^{l\,\dagger}\, V\, \boldsymbol{\psi}_m^l. \tag{4.133}$$

The $\boldsymbol{\psi}^l$ are solutions of (4.116), with $E = W - mc^2$,

$$c\boldsymbol{\sigma}\cdot\mathbf{p}\boldsymbol{\psi}^s + \mathbf{V}\boldsymbol{\psi}^l = E\boldsymbol{\psi}^l \tag{4.134}$$

and we see, remembering that $\boldsymbol{\psi}^s$ is of order $1/c$, that the integral is independent of c. This means that the downward shift of the electron eigenvalues caused by addition of an extra large component basis function is larger than the upward shift due to the inclusion of an extra small component basis function. If the large and small basis functions are not related by $\chi^s = \boldsymbol{\sigma}\cdot\mathbf{p}\chi^l$, then the integral in (4.128) does not vanish in general, both shifts scale independently of c and we cannot determine their relative magnitude.

In conclusion, a kinetically balanced basis set solves the problem of variational collapse and ensures that a variational scheme for the Dirac equation converges, from above, to the correct lowest energy eigenvalue.

4.3.4 The Relativistic Embedded Square Well

For comparison with the non-relativistic embedding method and to demonstrate a simple application we now look at the solution of the one-dimensional square well problem with the relativistic embedding method. We consider the same potential as before, V for $-a/2 < z < a/2$ and 0 otherwise. Once again, we can treat symmetric and antisymmetric solution separately and will focus on the symmetric solutions. To demonstrate that the energy eigenvalue, W, is minimised when $W = w$, where w is the energy at which we evaluate the embedding potential, we choose the trial solution inside the well to be

$$\chi(z) = \begin{pmatrix} \cos(kz)\phi_{m_s} \\ \mathrm{i}\gamma\sigma_z k \sin(kz)\phi_{m_s} \end{pmatrix}, \qquad \gamma = \frac{c\hbar}{w - V + mc^2}, \qquad k = \frac{1}{c\hbar}\sqrt{(w-V)^2 - (mc^2)^2}. \tag{4.135}$$

In the region outside the well, solutions satisfying suitable boundary conditions at $\pm\infty$ are the plane waves

$$\psi_<(z) = \begin{pmatrix} \phi_{m_s} \\ -\gamma_0\sigma_z k_0 \phi_{m_s} \end{pmatrix} \mathrm{e}^{-\mathrm{i}k_0 z}, \qquad \psi_>(z) = \begin{pmatrix} \phi_{m_s} \\ \gamma_0\sigma_z k_0 \phi_{m_s} \end{pmatrix} \mathrm{e}^{\mathrm{i}k_0 z} \tag{4.136}$$

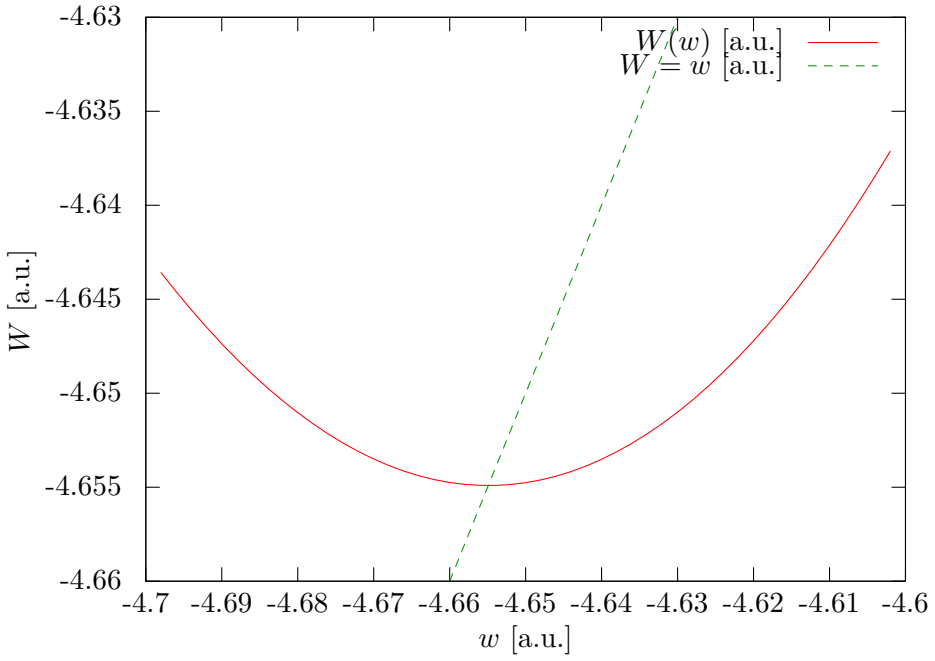

Figure 4-9: Variation of the energy expectation value, W, with the energy at which the embedding potential is evaluated, w, for a square well of depth $v = 5$ a.u. and width $a = \pi$ a.u. The minimum is at $w = -4.655$

where

$$\gamma_0 = \frac{c\hbar}{w + mc^2}, \qquad k_0 = \frac{1}{c\hbar}\sqrt{w^2 - (mc^2)^2}. \tag{4.137}$$

We can use these to calculate the embedding potential since it relates large and small components of the wavefunction at $z = \pm a/2$

$$\pm\gamma_0\sigma_z k_0\phi_{m_s}\mathrm{e}^{\pm\mathrm{i}k_0\frac{a}{2}} = \pm\mathrm{i}c\hbar\Gamma(w)\sigma_z\phi_{m_s}\mathrm{e}^{\pm\mathrm{i}k_0\frac{a}{2}} \tag{4.138}$$

and we obtain

$$\Gamma(w) = \frac{-\mathrm{i}}{c\hbar}\delta_{m_s m_s'}\sqrt{\frac{w - mc^2}{w + mc^2}}. \tag{4.139}$$

The factor of $\pm$ on the right hand side of (4.138) comes from the convention that the surface normal points out of the embedded region. We note that the embedding potential is diagonal in spin so we can consider each spin separately.

Substituting the trial function and embedding potential into the expression (4.86) for the expectation value, W, we find that is a minimum when $w = W$ (see figure 4-9). For a well of width $a = \pi$ a.u. and depth $V = 5$ a.u. the minimum is -4.655, which is the same as the non-relativistic value. This is not surprising since the first order correction to the kinetic energy of a relativistic particle scales as $1/c^4 \approx 3 \times 10^{-9}$ a.u.

We now expand the trial function in the well in a suitable, kinetically balanced

N	$w = V + mc^2$	$w = V/2 + mc^2$	exact
2	-4.6286881	-4.761651	-4.6360596
4	-4.6469680	-4.7297743	-4.6549433
6	-4.6481087	-4.7017345	-4.6548947
8	-4.6499066	-4.6699221	-4.6549309
10	-4.6551074	-4.6549379	-4.6548247

Table 4.2: Values of the lowest eigenvalue, W_1 calculated for different values of N and w for the well with $a = \pi$, $V = 5$ and $d = 1,26a$.

basis set. For symmetric solutions we choose

$$\chi(z) = \sum_{n=1}^{N} a_{l,n} \begin{pmatrix} \cos(k_n z) \\ 0 \end{pmatrix} + \sum_{n=1}^{N} a_{s,n} \begin{pmatrix} 0 \\ -\mathrm{i}\hbar k_n \sin(k_n z) \end{pmatrix}, \quad k_n = \frac{2(n-1)\pi}{d}, \quad n = 1, 2, 3, \cdots \tag{4.140}$$

where we have selected spin up since the problem is spin degenerate. The length $d > a$ is again chosen so that the basis functions do not satisfy any particular boundary conditions at $z = \pm a/2$. These basis functions are substituted into the expressions (4.107a)-(4.107f) for the matrix elements and the eigenvalues W_n calculated by solving the matrix equation

$$\sum_{n'=1}^{N} \begin{bmatrix} H_{ll} & H_{ls} \\ H_{sl} & H_{ss} \end{bmatrix}_{nn'} \begin{bmatrix} a_l \\ a_s \end{bmatrix}_{n'} = W_n \sum_{n'=1}^{N} \begin{bmatrix} O_{ll} & 0 \\ 0 & O_{ss} \end{bmatrix}_{nn'} \begin{bmatrix} a_l \\ a_s \end{bmatrix}_{n'}. \tag{4.141}$$

As in the non-relativistic case, we choose an input value, w, for the energy at which we evaluate the embedding potential and compute the eigenvalues. As we increase the size of the basis, for fixed w, we converge to the true eigenvalue, W_n from above. Table 4.2 shows some examples of this for various input w and sizes, N, of basis for a well of depth $V = -5$ a.u., width $a = \pi$ a.u. and length $d = 1.26a$ over which the basis functions are defined. For a given size of basis we can also find a particular eigenvalue, W_n, self consistently by choosing a starting value of w, calculating W_n and using this as an input w for a second cycle, continuing until $w = W_n$. The 'exact' values are calculated in this way. These results demonstrate that the relativistic embedding method is a variational method which converges to the eigenvalues of the system in exactly the same way as the non-relativistic case.

4.4 Self Consistent Embedding

At this point we have covered the basic ingredients necessary to perform *ab initio* relativistic embedded surface calculations. We have compared the single particle Dirac and Schrödinger equations and seen that the Dirac theory is preferable since it includes all of the relativistic effects necessary for dealing with heavy elements and the spin-orbit interaction which is particularly important for some interesting surface physics. Density Functional Theory provides a way of performing accurate calculations on large numbers of interacting electrons within a single particle framework for both relativistic and non-relativistic systems. We introduced the embedding method, which is an efficient way of treating systems which can be divided into distinct regions, and saw that this is a natural way of dealing with surfaces, particularly since it represents the bulk as a true semi-infinite crystal. This was then extended to incorporate relativistic effects via the Dirac embedding scheme. We now need to combine DFT and the embedding method to be able to perform fully relativistic surface electronic structure calculations. The remainder of this thesis describes how this has been achieved.

A typical non-relativistic surface embedding calculation is divided into three separate parts. First, a DFT calculation is performed to generate a bulk potential and this is then used to generate the embedding potential for the bulk. An analytic expression is typically used for the vacuum embedding potential. The second step is to generate a self consistent surface potential. To do this, the embedded Green's function is calculated for the embedded region and the charge density generated from this Green's function. From the charge density a new potential is constructed, the embedded Green's function is calculated to give a new charge density and the cycle is repeated to self-consistency. The last part of the calculation is to use the self-consistent surface potential to give a final embedded Green's function for the system, from which we can calculate charge densities, densities of states and so on.

We will end this chapter by introducing the basis functions for expanding the trial function in an embedded surface calculation.

4.5 The LAPW Basis

In the embedding method we have seen that if we express the trial function in some basis of suitable functions, we can calculate the Green's function by inverting the matrix,

$$[G(W)]_{nn'} = [WO - H + \Gamma]^{-1}_{nn'} \,. \tag{4.142}$$

which involves the Hamiltonian, H, overlap, O and any embedding potentials, Γ. There are some considerations which must be balanced when performing this calculation. Firstly, the computational time necessary for the inversion of an $N \times N$

matrix scales approximately as N^3, so the matrix must be as small as possible to increase computational efficiency. Secondly, the basis set must be capable of providing an accurate representation of the wavefunctions of the system. A third thing to bear in mind is storage and methods of sampling of the functions being used as a basis. Complicated functions will typically require more computational time to generate. These are general considerations that apply to all DFT calculations, where to solve the Kohn-Sham equations the wavefunctions are expanded in a set of basis functions.

An obvious first choice of basis is to use a simple plane wave expansion. The problem with this is that a large number of plane waves are needed to accurately represent the rapid spatial oscillations of atomic wavefunctions near the nucleus. In atoms, the lower energy, or core electrons are close to the nucleus and their wavefunctions decay rapidly with distance. This means that as the atoms are brought together to form a crystal, only the outer, valence electron wavefunctions will overlap and the core states remain almost completely unchanged from the isolated atom [78]. It is therefore reasonable to assume that we only need to consider variations in the valence wavefunctions and the core electrons provide a fixed potential in which the valence electrons move. This is the frozen core approximation [136]. A further improvement to reduce the size of the plane wave basis set is to remove the rapid oscillations of valence wavefunctions near the nucleus by the introduction of a pseudopotential.

The basic idea of a pseudopotential is that within some distance from the nucleus, the cut-off radius, r_C, the ionic Coulomb potential is replaced by a different potential which causes the wavefunctions to be smooth inside r_C. Outside of r_C the valence wavefunctions are identical to the true atomic states. A basic property of the atomic wavefunctions is that they are orthogonal to each other and it is this that causes the nodes of the atomic wavefunctions which leads to their rapid oscillation and norm-conserving pseudopotentials preserve the orthogonality of the valence states [137] [138]. In order to reproduce the scattering properties of the valence states, the pseudopotential depends on angular momentum and leads to non-local pseudopotentials [78]. Pseudopotential are calculated in two ways; from ab initio calculations on atoms or by fitting to experimental data and much of the ongoing work is to make them easily transferrable from one problem to another [78], [45], [139].

Pseudopotential reduce the number of plane waves needed for an accurate representation of atomic systems by concentrating only on the valence states. Another way to simplify the problem is to replace the plane wave expansion by an expansion in atomic wavefunctions in a region around the atom where rapid spatial oscillations are significant. in between atoms, where the wavefunctions are smooth, the plane wave expansion can be retained. This is the augmented plane wave approach (APW) [140]. The system is divided into atomic spheres, in which the wavefunction

is expanded in atomic solutions, and the interstitial region, where a plane wave basis is used. For close-packed materials, it is a reasonable approximation to assume that the potential inside the atomic sphere is spherically symmetric and constant in the interstitial region: the muffin tin approximation [141], [78].

The first step in an APW calculation is to solve for the atomic basis functions in the spheres [78]

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(r)\right]\psi(\mathbf{r};\epsilon) = \epsilon\psi(\mathbf{r}) \tag{4.143}$$

where ϵ is some chosen energy, called the pivot energy and $\psi(\mathbf{r})$ is

$$\psi(\mathbf{r};\epsilon) = \phi_{l,m_l}(\rho;\epsilon)Y_l^{m_l}(\hat{\boldsymbol{\rho}}). \tag{4.144}$$

Here, $Y_l^{m_l}(\hat{\boldsymbol{\rho}}$ are spherical harmonics, $\boldsymbol{\rho} = \mathbf{r} - \mathbf{R}_\alpha$, $\hat{\boldsymbol{\rho}} = (\mathbf{r} - \mathbf{R}_\alpha)/|\mathbf{r} - \mathbf{R}_\alpha|$ with $\mathbf{R}_\alpha$ the centre of the α^{th} atomic sphere and the radial functions depend on the pivot energy. In the interstitial, where the potential is constant, the wavefunctions are plane waves

$$\psi(\mathbf{r}) = \mathrm{e}^{\mathbf{i}\mathbf{k}\cdot\mathbf{r}}. \tag{4.145}$$

These basis functions are then matched in amplitude at the atomic sphere radii. The wavefunction is then expanded in these basis functions

$$\Psi(\mathbf{r}) = \sum_i a_i\psi_i(\mathbf{r}) \tag{4.146}$$

and leads to the usual matrix equation in the expansion coefficients, Hamiltonian and overlap

$$\sum_j [H]_{ij}\,[a]_j = \epsilon\sum_j [O]_{ij}\,[a]_j\,. \tag{4.147}$$

The problem is that the basis functions depend on the pivot energy, ϵ and the matrix equation becomes highly nonlinear and difficult to solve [45].

One solution to this problem of a nonlinear matrix equation is to reformulate the problem in terms of the Green's function. Rather than solving the matrix equation for the expansion coefficients of the wavefunction and using this to construct the charge density, the charge density and density of states, these can be calculated directly from the Green's function. This is the basis of modern Korringa-Kohn-Rostoker (KKR) calculations, based on a multiple scattering approach [92], [91]. The basic idea is that the Green's function, $G(\mathbf{r},\mathbf{r}')$ is a propagator, describing the propagation of a 'test' particle from $\mathbf{r}$ to $\mathbf{r}'$ and is a natural quantity for describing the scattering of Bloch states in a crystal [45]. The multiple scattering problem is formulated in terms of a scattering path operator, τ_{ij} which describes the scattering of a wave from potential i to potential j. The scattering path operator is related to the transition matrix, T, which enables the full Green's function to be

calculated from the single site Green's function via a Dyson equation [45]. A particular advantage of the KKR method is that it can be applied to inhomogeneous systems [95] and has been adapted to full-potential calculations beyond the muffin tin approximation [102], [96].

A second improvement on APW methods is to remove the energy dependence of the basis functions and hence the matrix equation. This is achieved by Linearised Augmented Plane Waves (LAPWs) [142]. This is the basis set we will use to perform embedded surface calculations in this thesis.

One difficulty with the APW method is that we must solve a nonlinear matrix equation for the pivot energies, ϵ, of our system. A further problem is that there may be atomic solutions which are zero on the atomic sphere boundary and the coefficients which match these on to the plane waves vary strongly with the pivot energy. This leads to numerical difficulties in calculating bands which lie close to these energies. However, although the atomic solution may be zero at the sphere baoundary, in general, its energy derivative will be non-zero [137]. The central idea of LAPW basis sets is that the basis functions inside the atomic spheres are an atomic solution plus its energy derivative, matched in amplitude and radial derivative on to plane waves in the interstitial region [142]. The main advantage of the method is that if the pivot energy ϵ differs from the actual band energy, E, then the errors in the wavefunction are of order $(E-\epsilon)^4$ [137], meaning that a given pivot energy is accurate over a reasonably large energy range. Pivot energies are typically chosen to lie at the centre of mass of the density of states [137].

The additional flexibility introduced by including the energy derivative of the atomic solutions means that they are applicable to full potential calculations since they can suitably treat non-spherical potentials in the atomic spheres [137]. However, the requirement that the functions match not only in amplitude but also in radial derivative at the atomic sphere boundaries means that they generally require a higher number of plane waves in the interstitial region [137].

Aside from their suitability for full potential calculations, LAPWs are chosen as the basis set for the calculations in this thesis for two reasons. Firstly and most importantly, many previous non-relativistic surface embedding calculations have been successfully performed using LAPWs, [106], [110], [46], [118], [143], and computer codes to perform these calculations already exist. Secondly, the extension of non-relativistic LAPWs to relativistic LAPWs is reasonably straightforward. Full potential LAPW calculations are commonly abbreviated FLAPW.

4.5.1 Non-Relativistic LAPW Basis Functions

In non-relativistic surface embedding calculations, the basis functions in the interstitial region are plane waves [106], [110]

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sqrt{\frac{2}{\mathbb{V}}} \mathrm{e}^{\mathrm{i}\mathbf{k}_{\parallel}\cdot\mathbf{r}_{\parallel}} \cos(k_n z) \tag{4.148}$$

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sqrt{\frac{2}{\mathbb{V}}} \mathrm{e}^{\mathrm{i}\mathbf{k}_{\parallel}\cdot\mathbf{r}_{\parallel}} \sin(k_n z) \tag{4.149}$$

where $\mathbb{V}$ is the volume of the unit cell, $\mathbf{k} = \mathbf{k}_{\parallel} + \mathbf{g} + k_n$ with $\mathbf{g}$ a two-dimensional reciprocal lattice vector, $\mathbf{k}_{\parallel}$ is in the first two-dimensional Brillouin zone and k_n is the wavevector in the direction normal to the surface. Basis functions in the α^{th} atomic sphere are given by

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{l\, m_l} \left[A^{\mathbf{k}}_{l\, m_l} u_l(\rho) + B^{\mathbf{k}}_{l\, m_l} \dot{u}_l(\rho) \right] Y_{l\, m_l}(\hat{\boldsymbol{\rho}}) \tag{4.150}$$

with $\rho = |\mathbf{r} - \mathbf{R}_\alpha|, \hat{\boldsymbol{\rho}} = (\mathbf{r} - \mathbf{R}_\alpha)/\rho$ where $\mathbf{R}_\alpha$ is the centre of the α^{th} atomic sphere. For calculations based on the Schrödinger equation, $u_l(\rho)$ is a solution of the radial Schrödinger equation at the angular momentum dependent pivot energy $\epsilon_{\alpha\, l}$ and $\dot{u}_l(\rho)$ is its energy derivative. The coefficients $A^{\mathbf{k}}_{l\, m_l}$ and $B^{\mathbf{k}}_{l\, m_l}$ are found by matching the radial solutions in amplitude and first radial derivative to the plane waves at the atomic sphere boundary, S_α. In scalar relativistic calculations, $u_l(\rho)$ is a two-component solution of the Koelling-Harmon equation at energy $\epsilon_{\alpha\, l}$ and $\dot{u}_l(\rho)$ its energy derivative. In order to allow matching of the two-component radial solutions to the one-component plane waves, $u_l(\rho)$ is chosen so that its small component vanishes at S_α [110]. The $A^{\mathbf{k}}_{l\, m_l}$ and $B^{\mathbf{k}}_{l\, m_l}$ are again found by matching the radial solutions in amplitude and first radial derivative to the plane waves at $\rho = S_\alpha$. Since we will not be needing expressions for the coefficients $A^{\mathbf{k}}_{l\, m_l}$ and $B^{\mathbf{k}}_{l\, m_l}$ in this work we will not provide them here but refer the reader to the article of Inglesfield and Benesh where they are listed [106].

In a real calculation a finite basis must be used and the set of plane waves is generated up to some maximum cut-off length of 2-d reciprocal lattice vector, g_{cut}. All g-vectors with $|\mathbf{g}| \leq g_{cut}$ are included and these are organised into sets which are equivalent under the symmetry operations which leave the lattice unchanged. These sets of symmetrically equivalent g-vectors are the stars of $\mathbf{k}$ [137]. The values of k_n must also be restricted to a finite set and this is done by defining them over some distance, $\mathbb{L}$, in the direction perpendicular to the surface (z), such that

$$k_n = \frac{n\pi}{\mathbb{L}}, \tag{4.151}$$

where n is even for plane wave basis functions which are cos-like and odd for sin-like, up to some maximum value N. As in the embedded square well, we choose $\mathbb{L}$ to be larger than the 'physical' size of the embedded region so that the plane waves do not satisfy any particular boundary conditions on the embedding surface, retaining variational freedom in the basis functions. The atomic sphere basis functions are calculated up to some angular momentum cut-off l_{max}, determined by the sphere radius S_α and plane wave cut-off g_{cut} [137]

$$l_{\max} = R_\alpha g_{cut}. \tag{4.152}$$

Typical values of l_{max} are in the region 8-10 [137], [106], [110].

To construct the matrix representation of the embedded Green's function at energy E we need to invert

$$\left[E0 - H + G_0^{-1}\right]_{\mathbf{k}\mathbf{k}'} \tag{4.153}$$

and matrix elements of the Hamiltonian, H, and overlap, O, are given by

$$[H]_{\mathbf{k}\mathbf{k}'} = \int_{\mathbb{V}} \mathrm{d}^3\mathbf{r}\, \psi^*_{\mathbf{k}}(\mathbf{r}) H \psi_{\mathbf{k}'}(\mathbf{r}) \tag{4.154}$$

$$\tag{4.155}$$

$$[O]_{\mathbf{k}\mathbf{k}'} = \int_{\mathbb{V}} \mathrm{d}^3\mathbf{r}\, \psi^*_{\mathbf{k}}(\mathbf{r}) \psi_{\mathbf{k}'}(\mathbf{r}). \tag{4.156}$$

Matrix elements of the embedding potential require special attention and will be discussed in the following section on the embedding potential. The Green's function is then constructed by an expansion in the LAPW basis functions

$$G(\mathbf{r}, \mathbf{r}'; E) = \sum_{\mathbf{k}\mathbf{k}'} G_{\mathbf{k}\mathbf{k}'}(E) \psi_{\mathbf{k}}(\mathbf{r}) \psi^*_{\mathbf{k}'}(\mathbf{r}). \tag{4.157}$$

Once again, these matrix elements will not be needed here and the reader is referred to the expressions in [106].

The wavefunctions of the system are expanded in plane waves in the interstitial region and spherical functions in atomic spheres, and the charge density is represented in a similar way. To simplify the expansion and increase computational efficiency, the symmetry of the lattice is exploited, leading to the stars of k used for the basis functions and lattice harmonics, $K_{\nu\alpha}(\hat{\boldsymbol{\rho}})$ in the atomic spheres. Lattice harmonics are linear combinations of spherical harmonics which are invariant under the symmetry operations of the lattice (the site symmetry)

$$K_{\nu\alpha}(\hat{\boldsymbol{\rho}}) = \sum_{m_l} c^\alpha_{\nu m_l} Y_{l\, m_l}(\hat{\boldsymbol{\rho}}). \tag{4.158}$$

The $\{c^{\alpha}_{\nu m_l}\}$ are expansion coefficients calculated from the rotational matrices of the site symmetry and the sum is only over m_l because rotations do not couple spherical harmonics with different l [137]. The charge density in the spheres is [110]

$$n(\boldsymbol{\rho}) = \sum_{\nu} n_{\nu\alpha}(\rho) K_{\nu\alpha}(\hat{\boldsymbol{\rho}}). \tag{4.159}$$

In the interstitial region, the charge density is expanded in plane waves

$$n(\mathbf{r}) = \sum_{\mathbf{g}\,n} n_{\mathbf{g}\,n} \mathrm{e}^{\mathrm{i}\mathbf{g}\cdot\mathbf{r}_{\parallel}} \cos(k_n z) \tag{4.160}$$

where the sum is over the two-dimensional reciprocal lattice vectors in the stars of k, $\mathbf{g}$ and the set of perpendicular wavevectors, k_n. The cos dependence is introduced to ensure that $n(\mathbf{r})$ is real [110].

4.5.2 Relativistic LAPWs

The extension of the LAPW basis functions to relativistic calculations is rather straightforward, but as usual there are some subtle considerations. Basis functions in the interstitial region are four component relativistic plane waves [144], [137]

$$\psi_{\mu}(\mathbf{r}) = \sqrt{\frac{2}{\mathbb{V}}} \left[\begin{pmatrix} \phi_{m_s} \\ \gamma_k \boldsymbol{\sigma}\cdot\mathbf{k}^{(+)}\phi_{m_s} \end{pmatrix} \mathrm{e}^{\mathrm{i}k_n z} \pm \begin{pmatrix} \phi_{m_s} \\ \gamma_k \boldsymbol{\sigma}\cdot\mathbf{k}^{(-)}\phi_{m_s} \end{pmatrix} \mathrm{e}^{-\mathrm{i}k_n z} \right] \mathrm{e}^{\mathrm{i}(\mathbf{k}_{\parallel}+\mathbf{g})\cdot\mathbf{r}_{\parallel}} \tag{4.161}$$

where the ϕ_{m_s} are Pauli 2-spinors and the index $\mu = (\mathbf{k},\, m_s)$ labels the wavevector and spin and

$$\mathbf{k}^{(\pm)} = \mathbf{k}_{\parallel} + \mathbf{g} \pm k_n \tag{4.162}$$

is introduced to give the large component a sin-like or cos-like form by analogy with the non-relativistic basis. The k_n and stars of k are identical to those in the non-relativistic case. The prefactor

$$\gamma_k = \frac{c\hbar}{W_k + mc^2}, \qquad W_k = +\sqrt{\hbar^2 c^2 k^2 + m^2 c^4}. \tag{4.163}$$

where the positive square root is taken because we are interested in electron-like solutions. The factor of $\boldsymbol{\sigma}\cdot\mathbf{k}$ in the small component means that these basis functions satisfy the kinetic balance condition.

The wavefunction in the atomic spheres is expanded in four component spherical solutions to the Dirac equation

$$\psi_{\mu}(\mathbf{r}) = \sum_{\Lambda} \left[A_{\Lambda\mu} \begin{pmatrix} g_{\kappa}(\rho)\Omega_{\Lambda}(\hat{\boldsymbol{\rho}}) \\ \mathrm{i} f_{\kappa}(\rho)\Omega_{\overline{\Lambda}}(\hat{\boldsymbol{\rho}}) \end{pmatrix} + B_{\Lambda\mu} \begin{pmatrix} \dot{g}_{\kappa}(\rho)\Omega_{\Lambda}(\hat{\boldsymbol{\rho}}) \\ \mathrm{i}\dot{f}_{\kappa}(\rho)\Omega_{\overline{\Lambda}}(\hat{\boldsymbol{\rho}}) \end{pmatrix} \right]. \tag{4.164}$$

The functions $g_{\kappa}(\rho)$ and $f_{\kappa}(\rho)$ are the usual solutions to the coupled radial equa-

tions (2.32) at pivot energy W_κ and $\dot{g}_\kappa(\rho)$ and $\dot{f}_\kappa(\rho)$ are their energy derivatives. The form of these spherical basis functions ensures that they satisfy the kinetic balance condition. The coefficients $A_{\Lambda\mu}$ and $B_{\Lambda\mu}$ are calculated by matching the spherical functions to the plane waves at the sphere boundary, S_α. Because the radial functions and plane waves are both four component spinors there is no requirement that the small components are zero at S_α. There are two possible sets of matching coefficients; we can match the amplitude of large and small components on S_α or we can match the amplitude and first derivative of the large component by analogy with the non-relativistic LAPWs. Both of these options will be investigated in the calculations presented in later chapters. Details of the determination of $A_{\Lambda\mu}$ and $B_{\Lambda\mu}$ are presented in appendix (A). The dependence of the angular momentum quantum numbers $\Lambda = (\kappa, \mu)$ on l (see (2.26)) means that l_{max} can be determined in the same way as before.

The radial functions $g_\kappa(\rho)$ and $f_\kappa(\rho)$ oscillate rapidly near the nucleus and become smoother as they approach S_α. To enable efficient sampling they are tabulated on radial grid points which are more closely spaced near the centre of the atomic sphere. A suitable coordinate is described by

$$\rho(x) = \mathrm{e}^x \tag{4.165}$$

and we define the functions

$$P_\kappa(x) = \rho g_\kappa(\rho), \qquad Q_\kappa(x) = \rho f_\kappa(\rho) \tag{4.166}$$

normalised over the sphere such that

$$a_\kappa^2 \int_{S_\alpha(x)} \mathrm{d}x\, \rho(x) \left[P_\kappa^2(x) + Q_\kappa^2(x)\right] = 1, \tag{4.167}$$

so that

$$a_\kappa = \frac{1}{\sqrt{\int_{S(x)} \mathrm{d}x\, \rho(x) \left[P_\kappa^2(x) + Q_\kappa^2(x)\right]}} \tag{4.168}$$

and

$$\dot{a}_\kappa = -a_\kappa^3 \int_{S(x)} \mathrm{d}x\, \rho(x) \left[P_\kappa(x)\dot{P}_\kappa(x) + Q_\kappa(x)\dot{Q}_\kappa(x)\right] \tag{4.169}$$

Matrix elements are calculated in a similar way to the non-relativistic LAPWs,

except we now have Hermitian conjugation in the place of complex conjugation

$$[H]_{\mu\mu'} = \int_{\mathbb{V}} \mathrm{d}^3\mathbf{r}\, \psi^\dagger_\mu(\mathbf{r}) H \psi_{\mu'}(\mathbf{r}) \tag{4.170}$$

$$\tag{4.171}$$

$$[O]_{\mu\mu'} = \int_{\mathbb{V}} \mathrm{d}^3\mathbf{r}\, \psi^\dagger_\mu(\mathbf{r}) \psi_{\mu'}(\mathbf{r}). \tag{4.172}$$

The Green's function at energy W is expanded in the double basis

$$G(\mathbf{r}, \mathbf{r}'; W) = \sum_{\mu\mu'} \psi_\mu(\mathbf{r}) G_{\mu\mu'}(W) \psi^\dagger_{\mu'}(\mathbf{r}). \tag{4.173}$$

with the matrix $G_{\mu\mu'}(W)$ calculated by inverting

$$[W0 - H + \Gamma]_{\mu\mu'}\,. \tag{4.174}$$

where the embedding potential matrix element is now denoted by Γ. Detailed expressions for these matrix elements used for the calculations described later on are presented in appendix (A). The composite index $\mu = (\mathbf{k}, m_s)$.

The charge density can be expanded in exactly the same way as in the non-relativistic case since it is a scalar quantity, not a four-component spinor.

Chapter 5

THE TRANSFER MATRIX, COMPLEX BAND STRUCTURES AND THE EMBEDDING POTENTIAL

In the preceding chapter, we mentioned that a self-consistent surface embedding calculation was essentially divided into three parts; the bulk potential, the embedding potentials for the bulk and vacuum regions and the self-consistent calculation in the embedded surface region. The bulk calculation can be performed using any of the standard methods of DFT mentioned previously, although if the embedding part of the calculation is implemented in for example, the LAPW basis, it makes sense to calculate the bulk potential in the same basis. Once we have the bulk potential, we then need to somehow use it to generate the embedding potential that replaces the substrate in the surface calculation. That is the focus of this chapter and we will see that there are several ways we can do this, although one is preferable for a number of reasons. For example it makes available the bulk complex band structure, which is particularly useful for transport calculations, as a by-product of generating the embedding potential.

5.1 The Embedding Potential

The embedding potential is the key quantity in the embedding method. It is an energy dependent potential, defined over the embedding surface S, which ensures that the trial function in the embedded region matches correctly on to the solution in the substrate. In general, the embedding surface S may be complicated and for a surface calculation the natural choice of S on the bulk side is one which weaves around the atomic spheres (see figure 5-1). The problem with this is that a surface expansion of

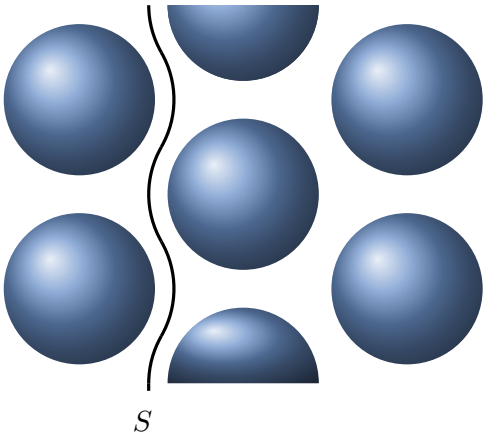

Figure 5-1: An embedding surface, S, which weaves between atomic spheres.

the embedding potential on a curvy surface is not suitable since it requires either a large number of plane waves or a complicated set of surface expansion functions and makes evaluation of matrix elements computationally demanding. An ideal choice would be to find a suitable planar surface which would enable a simple expansion of the embedding potential. It is possible to shrink the atomic spheres so that such a plane can be inserted between adjacent layers, but at the cost of a significant increase in the number of interstitial plane waves which is not desirable. In addition, this does not change the nature of the wavefunction which will have string spatial variation on the plane requiring a large number of basis functions for an accurate expansion. Another option is to define a planar surface which cuts through the atomic spheres and making a dual basis expansion but this leads to problems with including the sphere 'caps' [106], [105] [46]. Experience has shown that the best solution is to add an auxiliary 'buffer' volume between the curvy surface S and a planar surface on which the embedding potential is expanded [106], [105], [110], [89], [46].

5.1.1 Buffer Volumes

In non-relativistic embedding, the embedding potential ensures that the wavefunction, ϕ, in region I, the embedded region, has the correct amplitude and normal derivative on the embedding surface S so that it matches smoothly on to the wavefunction, ψ in region II (see section 4.2). In the literature, this matching of amplitude and normal derivative of ψ and ϕ on S is commonly referred to by saying they have the same *logarithmic derivative* (see for example [105], [106]).

To transfer the embedding surface from the curvy S to the plane P we add an auxiliary, buffer region Δ, between S and P, shown in figure 5-2. When the trial function is converged, it is a solution of the Schrödinger (or Dirac) equation over the whole system and the embedding potential ensures it has the correct logarithmic derivative on S. This means that the boundary condition on S can be transferred

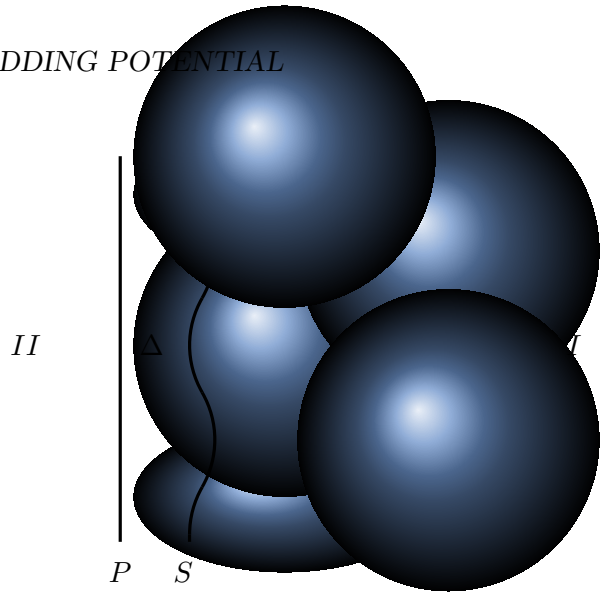

Figure 5-2: The embedding surface can be transferred from the curved surface S to a plane P via the addition of an auxiliary 'buffer' region, Δ.

to P by integrating through the buffer region Δ [106], [105]. The curved surface, S, is in the interstitial region so we expect that the amplitude and derivative of the wavefunction to be relatively smooth and suggests that we can use a plane wave expansion of the embedding potential matrix elements on the plane P. Plane waves also describe free particles and we can therefore choose a constant potential in Δ. Thus, by adding the buffer region Δ with constant potential, we are able to use a simple plane wave expansion of the embedding potential on a plane.

5.1.2 Methods for the Generating the Non-Relativistic Bulk Embedding Potential

Adding a buffer region to transfer the boundary condition to a plane enables us to construct simple matrix elements for the embedding potential. If we take a 'representative layer' of atoms and add buffer regions to both edges, as shown in figure 5-3, it is possible to build up the embedding potential for the bulk half-space on a layer-by-layer basis. Non-relativistic embedding potentials, G_L^{-1} and G_R^{-1}, on the planes P_L and P_R can be expanded as a Fourier series in $\mathbf{k}_\parallel$ to give the $\mathbf{k}_\parallel$-resolved embedding potential, $G_0^{-1}(\mathbf{g}, \mathbf{g}'; \mathbf{k}_\parallel)$ [106]

$$G_0^{-1}(\mathbf{r}_\parallel, \mathbf{r}_\parallel{}'; \mathbf{k}_\parallel) = \frac{1}{\mathbb{A}} \sum_{\mathbf{g},\mathbf{g}'} G_0^{-1}(\mathbf{g}, \mathbf{g}'; \mathbf{k}_\parallel) \mathrm{e}^{\mathrm{i}(\mathbf{k}_\parallel + \mathbf{g})\cdot \mathbf{r}_\parallel} \mathrm{e}^{-\mathrm{i}(\mathbf{k}_\parallel + \mathbf{g}')\cdot \mathbf{r}_\parallel{}'}. \tag{5.1}$$

In the muffin tin approximation, the potential in the interstitial is constant and by inverting the Green's function for plane waves incident on the embedding planes, the $\mathbf{k}_\parallel$-resolved embedding potential can be written in terms of the reflection matrix, $R_{\mathbf{g},\mathbf{g}'}$ for a semi infinite crystal [106], [105]

$$G_0^{-1}(\mathbf{g}, \mathbf{g}'; \mathbf{k}_\parallel) = -\frac{\mathrm{i}k_{\mathbf{g},z}}{2} \left[(1 - R)(1 + R)\right]_{\mathbf{g},\mathbf{g}'}, \qquad k_{\mathbf{g},z} = \sqrt{2E - |\mathbf{g} + \mathbf{k}_\parallel|^2}. \tag{5.2}$$

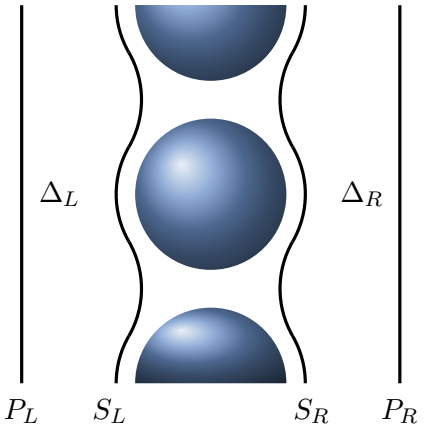

Figure 5-3: Geometry for generating a bulk embedding potential by adding buffer regions, Δ_L and Δ_R, to a representative layer of the bulk crystal.

The reflection coefficient can be determined by Layer-KKR methods; the reflection matrix for the semi-infinite crystal being built up by considering layer by layer scattering [99]. However, if we wish to use the embedding potential generated from this LKKR reflection matrix in a full-potential LAPW surface calculation, the fact that the substrate is treated in the muffin tin approximation will limit the applicability of the method to close-packed materials [106], [105].

To construct an embedding potential suitable for full-potential calculations, Crampin et. al. developed an iterative procedure for building the embedding potential layer-by-layer (see [105] and [109] for technical details). The basic idea is that by starting with an initial guess for G_0^{-1}, the Green's function for the embedded layer is constructed with zero embedding potential on P_R. This means that the Green's function has zero normal derivative on P_R and can be inverted on P_R to give the embedding potential on P_R (see section 4.2). This is used as a new input embedding potential on P_L and the procedure is repeated to give a new embedding potential on P_R. Repeating this process grows the embedding potential on P_R layer-by-layer and can be repeated to self-consistency, so that the embedding potential on P_R is the same as the input embedding potential on P_L to within a desired level of accuracy. At this point, the embedding potential will be that of the bulk crystal. In these embedding calculations, the plane surface P is typically chosen to lie half way between atomic layers, to minimise the distance over which the boundary condition must be transferred [105]. Because the buffer region is unphysical, the same buffer region must subsequently be used in the surface part of the calculation so that it can be 'subtracted off'. This means we still need to know the form of the curved embedding surface S [110], [118].

An improvement to these iterative methods was introduced by Ishida [110]. Previously, the specific form of the curvy surface S was needed when transferring the boundary condition to the plane P. The addition of buffer regions Δ_L and Δ_R on

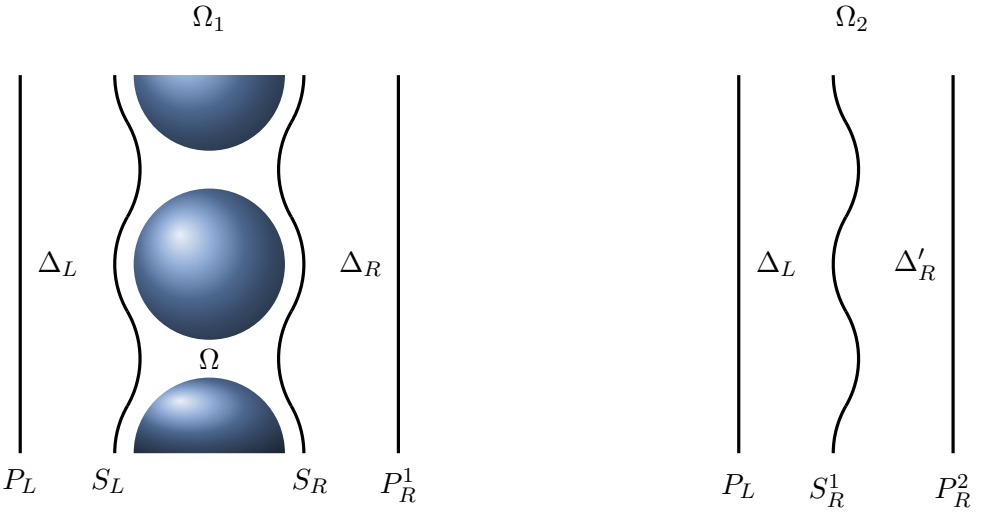

Figure 5-4: Setup for generating the bulk embedding potential without the need to specify the curved surfaces S_L and S_R.

both sides of the representative layer, Ω, (the 'physical' embedded region) creates the region Ω_1, shown in figure 5-4. However, these can be 'subtracted off' from the embedding potential calculation by introducing the additional region Ω_2 consisting of Δ_L and Δ'_R. If successive layers in the bulk crystal are related by a translation, $\mathbf{d}$, then Ω_2 is constructed from Δ_L and $\Delta'_R = \Delta_R - \mathbf{d}$, i.e. region Ω_1 with Ω removed. We are still free to choose a constant potential in the buffer regions to allow plane wave matrix elements. The iterative procedure is modified as follows. An initial 'guess' for $G^{-1}_{P^2_R}$ is chosen on P^2_R and the Green's function for Ω_2 with zero amplitude normal derivative on P_L is calculated, by setting the embedding potential on P_L to zero, and inverted to obtain the embedding potential, $G^{-1}_{P_L}$ on P_L. Taking $G^{-1}_{P_L}$ as an input on P_L we then find the Green's function for Ω_1 with zero amplitude normal derivative on P^1_R by setting the embedding potential on P^1_R to zero and invert this to get the embedding potential $G^{-1}_{P^1_R}$ on P^1_R. The self-consistency condition is then $G^{-1}_{P^2_R} = G^{-1}_{P^1_R}$. Because the potential is the same in Δ_R and Δ'_R, this is equivalent to $G^{-1}_{S_L} = G^{-1}_{S^1_R}$. Since Δ_L appears in both Ω_1 and Ω_2 we have $G^{-1}_{S_L} = G^{-1}_{S_R}$ and this is the self-consistency condition on the 'physical' curved embedding surfaces.

The iterative methods for generating the bulk embedding potential have certain disadvantages. although they generally require only a few iterations to converge, this number can increase over certain energy ranges when the nature of the states changes rapidly with energy and especially when the imaginary part of the energy is small [105], [110], [118]. They have therefore been superceded by an alternative method of generating an embedding potential for the bulk substrate based on the transfer matrix [118]. This is the method we have developed to construct embedding potentials for relativistic calculations and which also enables the bulk complex band structureto be calculated.

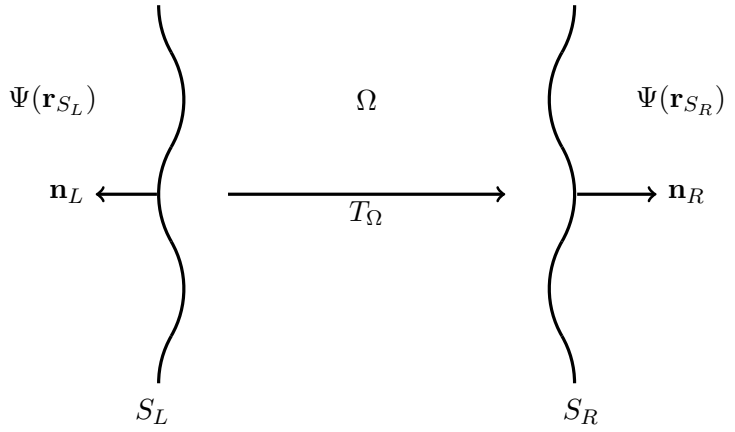

Figure 5-5: The transfer matrix, T_Ω, has the effect of transferring the boundary condition of the wavefunction Ψ from surface S_L to surface S_R, through the volume Ω. Surface normals, $\mathbf{n}_S$, are defined to point out of Ω.

5.2 The Transfer Matrix

In the context of non-relativistic embedding, the transfer matrix, t_Ω, relates the amplitude of the wavefunction, ψ and its normal derivative, $\partial\psi/\partial n_S$ on some surface, S_L, to these values on some other surface S_R, where S_L and S_R enclose some volume Ω as shown in figure 5-5 [118]. Hence

$$\Psi(\mathbf{r}_{S_R}) = \int_{S_L} \mathrm{d}^2\mathbf{r}'_{S_L} t_\Omega(\mathbf{r}_{S_R}, \mathbf{r}'_{S_L})\Psi(\mathbf{r}'_{S_L}) \tag{5.3}$$

where

$$\Psi(\mathbf{r}_S) = \begin{pmatrix} \psi(\mathbf{r}_S) \\ \\ \dfrac{\partial\psi(\mathbf{r}_S)}{\partial n_S} \end{pmatrix}. \tag{5.4}$$

The definition (5.3) is usually written in the more transparent, simplified notation

$$\begin{pmatrix} \psi(\mathbf{r}_{S_R}) \\ \\ \dfrac{\partial\psi(\mathbf{r}_{S_R})}{\partial n_R} \end{pmatrix} = \begin{pmatrix} t_\Omega^{11} & t_\Omega^{12} \\ \\ t_\Omega^{21} & t_\Omega^{22} \end{pmatrix} \begin{pmatrix} \psi(\mathbf{r}_{S_L}) \\ \\ \dfrac{\partial\psi(\mathbf{r}_{S_L})}{\partial n_L} \end{pmatrix} \tag{5.5}$$

where the integrals over the appropriate surface are implied. The transfer matrix therefore transfers the boundary condition of the wavefunction from S_L to S_R. In order to incorporate the transfer matrix into embedding calculations we define surface normals, $\mathbf{n}_S$ to point *out of* Ω, in line with the standard convention.

This definition of the transfer matrix is easily extended to relativistic embedding

problems by defining

$$\begin{pmatrix} \psi_l(\mathbf{r}_{S_R}) \\ \psi_s(\mathbf{r}_{S_R}) \end{pmatrix} = \begin{pmatrix} T_\Omega^{ll} & T_\Omega^{ls} \\ T_\Omega^{sl} & T_\Omega^{ss} \end{pmatrix} \begin{pmatrix} \psi_l(\mathbf{r}_{S_L}) \\ \psi_s(\mathbf{r}_{S_L}) \end{pmatrix} \tag{5.6}$$

where the surface integrals are implied and each block of T is a 2×2, meaning that T is a 4×4 matrix which relates the amplitude of the large and small components on S_L and S_R. Equation (4.82) shows that the Green's function relates the value of the wavefunction on S to its amplitude in the volume bounded by S

$$\psi(\mathbf{r}) = \mathrm{i}c\hbar \int_S \mathrm{d}^2\mathbf{r}_S \cdot G(\mathbf{r}, \mathbf{r}_S)\boldsymbol{\alpha}\psi(\mathbf{r}_S) \tag{5.7}$$

For the geometry in figure (5-5), $S = S_L \cup S_R$. Letting $\mathbf{r} \to \mathbf{r}_{S_R}$ or S_L from within Ω and expanding the wavefunctions and Green's functions in large and small components gives, for the large component of ψ,

$$\begin{aligned}\psi_l(\mathbf{r}_R) = -\mathrm{i}c\hbar \int_{S_L} \mathrm{d}^2\mathbf{r}_L G_{ll}(\mathbf{r}_R, \mathbf{r}_L)\sigma_L\psi_s(\mathbf{r}_L) - \mathrm{i}c\hbar \int_{S_L} \mathrm{d}^2 r_L G_{ls}(\mathbf{r}_R, \mathbf{r}_L)\sigma_L\psi_l(\mathbf{r}_L) \\ + \mathrm{i}c\hbar \int_{S_R} \mathrm{d}^2\mathbf{r}'_R G_{ll}(\mathbf{r}_R, \mathbf{r}'_R)\sigma_R\psi_s(\mathbf{r}'_R) + \mathrm{i}c\hbar \int_{S_R} \mathrm{d}^2\mathbf{r}_R G_{ls}(\mathbf{r}_R, \mathbf{r}'_R)\sigma_R\psi_l(\mathbf{r}'_R)\end{aligned} \tag{5.8}$$

$$\begin{aligned}\psi_l(\mathbf{r}_L) = -\mathrm{i}c\hbar \int_{S_L} \mathrm{d}^2\mathbf{r}_L G_{ll}(\mathbf{r}_L, \mathbf{r}'_L)\sigma_L\psi_s(\mathbf{r}'_L) - \mathrm{i}c\hbar \int_{S_L} \mathrm{d}^2\mathbf{r}_L G_{ls}(\mathbf{r}_L, \mathbf{r}'_L)\sigma_L\psi_l(\mathbf{r}'_L) \\ + \mathrm{i}c\hbar \int_{S_R} \mathrm{d}^2 r_R G_{ll}(\mathbf{r}_L, \mathbf{r}_R)\sigma_R\psi_s(\mathbf{r}_R) + \mathrm{i}c\hbar \int_{S_R} \mathrm{d}^2 r_R G_{ls}(\mathbf{r}_L, \mathbf{r}_R)\sigma_R\psi_l(\mathbf{r}_R)\end{aligned} \tag{5.9}$$

where we have introduced the notation $\mathbf{r}_L = \mathbf{r}$ on S_L and $\sigma_L = \boldsymbol{\sigma} \cdot \mathbf{n}_L$ etc. These relationships indicate that the transfer matrix can be written in terms of the Green's function. To simplify the process we make use the fact that we are free to choose a Green's function which has vanishing small components on S_L and S_R. This freedom follows for the fact that the Green's function for the Hamiltonian H is defined

$$[H - W]\, G(W) = -\delta(\mathbf{r} - \mathbf{r}'), \qquad \mathbf{r}, \mathbf{r}' \in \Omega. \tag{5.10}$$

Ψ is an eigenfunction of H at energy W so that $[H - W]\,\Psi = 0$, hence

$$[H - W]\,(G(W) + \Psi) = -\delta(\mathbf{r} - \mathbf{r}') \tag{5.11}$$

and the definition of the Green's function is therefore not unique. By constructing the Green's function for Ω such that its small components vanish on S_L and S_R, i.e.

only G_{ll} is non-zero, the above expressions become

$$\psi_L^l = F_{LR}\psi_R^s + F_{LL}\psi_L^s, \qquad \psi_R^l = F_{RR}\psi_R^s + F_{RL}\psi_L^s. \tag{5.12}$$

We have introduced a much simplified notation where ψ_L^l is the large component of $\psi(\mathbf{r})$ on S_L,

$$F_{LR} = \mathrm{i}c\hbar G_{ll}(\mathbf{r}_L, \mathbf{r}_R)\boldsymbol{\sigma}\cdot\mathbf{n}_R, \tag{5.13}$$

etc. and multiplication implies integration over the appropriate surface. Rearranging (5.12) gives the following expression for T_Ω

$$T_\Omega = \begin{pmatrix} F_{RR}F_{LR}^{-1} & F_{RL} - F_{RR}F_{LR}^{-1}F_{LL} \\ \\ F_{LR}^{-1} & -F_{LR}^{-1}F_{LL} \end{pmatrix}. \tag{5.14}$$

The transfer matrix, T_Ω can therefore be constructed from the Green's function for region Ω with zero-amplitude small component on S

In practice, we obtain the Green's function in (5.14) from an embedding calculation. The embedded Green's function for region Ω is in general given by

$$G = [WO - H + \Gamma_L + \Gamma_R]^{-1} \tag{5.15}$$

where Γ_L and Γ_R are the embedding potentials on S_L and S_R. Recall that the embedding potential, Γ ensures that the solution of the embedded problem satisfies specific boundary conditions on the bounding surfaces, in particular relating the amplitude of small and large components (4.83). Putting $\Gamma_L = \Gamma_R = 0$ in (5.15) ensures that the Green's function has zero amplitude small component on S_L and S_R. It should be noted however that this Green's function is. in general, not physically meaningful and cannot be used to describe the electronic structure of the embedded region, but has been introduced to enable the calculation of T_Ω. We will now compare the relativistic and non-relativistic transfer matrices.

5.2.1 Non-Relativistic Limits of the Transfer Matrix

Having described a scheme for constructing the relativistic transfer matrix, it is instructive compare it with the corresponding non-relativistic transfer matrix, t_Ω, which relates the amplitude and normal derivative of the wavefunction on S_L and S_R and can be constructed from the Green's function for region Ω with zero normal derivative on S_L and S_R. This Green's function can be calculated from an embedding calculation where embedding potentials on S_L and S_R are set to zero. In a similar

notation to that introduced for T_Ω, the expression for t_Ω in terms of G is [118]

$$t_\Omega = \begin{pmatrix} G_{RR}G_{LR}^{-1} & -\frac{\hbar^2}{2m}G_{RL} + \frac{\hbar^2}{2m}G_{RR}G_{LR}^{-1}G_{LL} \\ -\frac{2m}{\hbar^2}G_{LR}^{-1} & -G_{LR}^{-1}G_{LL} \end{pmatrix}. \tag{5.16}$$

Writing out explicitly the quantities F_{LL} etc in (5.14), the relativistic transfer matrix is

$$T_\Omega = \begin{pmatrix} G_{RR}^{ll}\left[G_{LR}^{ll}\right]^{-1} & -\mathrm{i}c\hbar G_{RL}\sigma_L + \mathrm{i}c\hbar G_{RR}^{ll}\left[G_{LR}^{ll}\right]^{-1}G_{LL}^{ll}\sigma_L \\ -\frac{1}{\mathrm{i}c\hbar}\sigma_R\left[G_{LR}^{ll}\right]^{-1} & -\sigma_R\left[G_{LR}^{ll}\right]^{-1}G_{LL}^{ll}\sigma_L \end{pmatrix}, \tag{5.17}$$

where we have used the fact that $\sigma_S\sigma_S = \mathbb{I}_2$. From section 2.2.2 we have seen that the large-large component of the relativistic Green's function, which is the only component appearing in (5.17), reduces to the non-relativistic Green's function in the $c \to \infty$ limit and the expression for t_Ω and T_Ω are formally very similar. In the non-relativistic limit the diagonal blocks of T_Ω are identical to those of t_Ω. There are however problems with the off-diagonal blocks, in particular T_Ω^{ls} is proportional to c and blows up as $c \to \infty$, and T_Ω^{sl} varies as $1/c$ and vanishes.

We can regularize the non-relativistic limit by introducing the transformation

$$\tilde{\psi}_S(\mathbf{r}_S) = C_S\psi_S(\mathbf{r}_S), \qquad \tilde{T}_\Omega = C_R T_\Omega C_L^{-1} \tag{5.18}$$

with

$$C_S(\mathbf{r}_S, \mathbf{r}'_S) = \begin{pmatrix} \mathbb{I}_2 & 0 \\ 0 & \mathrm{i}\left[\frac{W - V(\mathbf{r}_S) + mc^2}{c\hbar}\right]\sigma_S(\mathbf{r}_S) \end{pmatrix}\delta(\mathbf{r}_S - \mathbf{r}'_S) \tag{5.19}$$

for $S = L$ or R. Under this transformation $\tilde{\psi}_R = \tilde{T}_\Omega\tilde{\psi}_L$ and as $c \to \infty$

$$\tilde{T}_\Omega \to \begin{pmatrix} G_{RR}G_{LR}^{-1} & -\frac{\hbar^2}{2m}G_{RL} + \frac{\hbar^2}{2m}G_{RR}G_{LR}^{-1}G_{LL} \\ -\frac{2m}{\hbar^2}G_{LR}^{-1} & -G_{LR}^{-1}G_{LL} \end{pmatrix} \tag{5.20}$$

where each entry is a 2×2 matrix. This now has an identical to form to the non-relativistic transfer matrix. Recall that the non-relativistic transfer matrix is defined by using the amplitude and normal derivative of the wavefunction, $\Psi(\mathbf{r}_S)$ and in the relativistic case by the large and small components on S. Therefore the transfer

matrix relates surface values of the quantities

$$\chi_{non-rel}(\mathbf{r}_S) = \begin{pmatrix} \psi(\mathbf{r}_S) \\ \\ \dfrac{\partial\psi(\mathbf{r}_S)}{\partial n_S} \end{pmatrix}, \qquad \chi_{rel}(\mathbf{r}_S) = \begin{pmatrix} \psi_l(\mathbf{r}_S) \\ \psi_s(\mathbf{r}_S) \end{pmatrix} = \begin{pmatrix} \psi_l(\mathbf{r}_S) \\ \\ \dfrac{-\mathrm{i}c\hbar}{W - V(\mathbf{r}_S) + mc^2}\boldsymbol{\sigma}\cdot\nabla\psi_l(\mathbf{r}_S) \end{pmatrix} \tag{5.21}$$

where we have used the general relationship between the large and small components in the relativistic case. Applying the transformation C_S to the relativistic wavefunction we see that, in the $c \to \infty$ limit, the transfer matrix T_Ω relates surface values of

$$\tilde{\chi}_{rel}(\mathbf{r}_S) = \begin{pmatrix} \psi_l(\mathbf{r}_S) \\ \\ \dfrac{\partial\psi_l(\mathbf{r}_S)}{\partial n_S} + \mathrm{i}\boldsymbol{\sigma}\cdot[\mathbf{n}_S \times \nabla\psi_l(\mathbf{r}_S)] \end{pmatrix} \tag{5.22}$$

Since the non-relativistic limit of the large component of the relativistic wavefunction is the non-relativistic wavefunction and is diagonal in spin, this is the vector used to define the non-relativistic transfer matrix, except for the second term in the lower component which we encountered before in section 4.3.1. This term arises from a coupling of the spin and momentum of a relativistic electron when the spin is not quantised along the electron's direction of motion. Note that if the surface S lies in the x-y plane and the electron described by χ is moving in the z-direction ($\mathbf{k}_\parallel = 0$), then the extra term vanishes.

5.2.2 An Analytic Example

To illustrate the general properties of the transfer matrix we will look at a simple example. We consider a region Ω of constant potential V between parallel planes S_L and S_R at $z = L$ and $z = R$. The transfer matrix for this system could, for example, be applied to scattering from a piecewise constant potential, a model which is used to study quantum well structures [54], [145].

The system is translationally invariant in the x-y plane and we can write the wavevector as $\mathbf{k} = \mathbf{k}_\parallel + k_z\hat{\mathbf{z}}$ where $\mathbf{k}_\parallel = (k_x, k_y)$ is the momentum parallel to S_L and S_R. To construct the $\mathbf{k}_\parallel$-resolved transfer matrix, $T_{\Omega,\mathbf{k}_\parallel}$, we need the Green's function in Ω whose small components vanish at $z = L$ and $z = R$. To get this we introduce the wavefunctions at energy W in Ω which have vanishing small component at $z = 0$

$$\psi_\nu(z) = \begin{pmatrix} \dfrac{c\hbar(\boldsymbol{\sigma}_{||}\cdot\mathbf{k}_\parallel + \sigma_z k_z)}{W - V - mc^2}\cos k_z z\,\phi_{m_s} \\ \\ \sin k_z z\,\phi_{m_s} \end{pmatrix}, \qquad k_z = \frac{1}{c\hbar}\sqrt{(W-V)^2 - (mc^2)^2 - c^2\hbar^2|\mathbf{k}_\parallel|^2} \tag{5.23}$$

where $\nu = (\mathbf{k}_\parallel, m_s)$[‡]. The Green's function with zero amplitude small component on S_L and S_R can then be constructed via the direct method, which yields

$$G_{\mathbf{k}_\parallel}(z, z', ; W) = \begin{cases} \dfrac{W - V - mc^2}{c^2\hbar^2 k_x \sin k_z d} \psi_{\mathbf{k}_\parallel, m_s}(z - L)\psi^{\times}_{\mathbf{k}_\parallel, m_s'}(z - R) & z < z' \\ \\ \dfrac{W - V - mc^2}{c^2\hbar^2 k_x \sin k_z d} \psi_{\mathbf{k}_\parallel, m_s}(z - R)\psi^{\times}_{\mathbf{k}_\parallel, m_s'}(z - L) & z > z' \end{cases} \tag{5.24}$$

where $d = R - L$. Letting z and z' approach the surface planes from within Ω gives

$$F_{LL} = \frac{\mathrm{i}c\hbar}{W - V - mc^2} \begin{pmatrix} k_z \cot k_z d & -\mathrm{i}(k_x - \mathrm{i}k_y) \\ -\mathrm{i}(k_x - \mathrm{i}k_y) & -k_z \cot k_z d \end{pmatrix}$$

$$F_{LR} = \frac{\mathrm{i}c\hbar}{W - V - mc^2} \begin{pmatrix} -k_z & 0 \\ 0 & k_z \end{pmatrix}$$

$$F_{RL} = \frac{\mathrm{i}c\hbar}{W - V - mc^2} \begin{pmatrix} k_z & 0 \\ 0 & -k_z \end{pmatrix}$$

$$F_{RR} = \frac{\mathrm{i}c\hbar}{W - V - mc^2} \begin{pmatrix} -k_z \cot k_z d & -\mathrm{i}(k_x - \mathrm{i}k_y) \\ -\mathrm{i}(k_x - \mathrm{i}k_y) & k_z \cot k_z d \end{pmatrix}, \tag{5.25}$$

and from these we get the transfer matrix

$$T_{\Omega,\mathbf{k}_\parallel} = \begin{pmatrix} \cos k_z d\, \mathbb{I}_2 + \dfrac{\mathrm{i}}{k_z}\boldsymbol{\sigma}_{||} \cdot \mathbf{k}_\parallel\, \sigma_z \sin k_z d & \dfrac{\mathrm{i}(W - V + mc^2)}{c\hbar k_z}\, \sigma_z \sin k_z d \\ \\ \dfrac{\mathrm{i}(W - V - mc^2)}{c\hbar k_z}\, \sigma_z \sin k_z d & \cos k_z d\, \mathbb{I}_2 + \dfrac{\mathrm{i}}{k_z}\boldsymbol{\sigma}_{||} \cdot \mathbf{k}_\parallel\, \sigma_z \sin k_z d \end{pmatrix}. \tag{5.26}$$

The diagonal blocks of $T_{\Omega,\mathbf{k}_\parallel}$ each contain a term proportional to $\boldsymbol{\sigma} \cdot \mathbf{k}_\parallel$, which are non-zero when the momentum is not in the direction of the spin. These terms are due to the spin-momentum coupling discussed above when comparing the non-relativistic limit of the boundary values of ψ. This means that we can only directly compare the relativistic and non-relativistic transfer matrices at $\mathbf{k}_\parallel = 0$, when the momentum is along the spin-quantisation axis. To take the non-relativistic limit of $T_{\Omega,\mathbf{k}_\parallel}$ we make the transformation (5.19) which regularizes the limit. For our model

[‡]Although they appear different to the plane wave solutions to the Dirac equation that we have seen before, these wavefunctions do describe electron states because the large (upper) component is $1/c$ times the small (lower component). We can see this by subtracting the rest energy from $W = E + mc^2$ and noting that the upper component is of order c and the lower component is order 1.

system, the transformation matrices are

$$C_R = \begin{pmatrix} \mathbb{I}_2 & 0 \\ 0 & \dfrac{\mathrm{i}}{\gamma_+}\sigma_z \end{pmatrix}, \quad C_L^{-1} = \begin{pmatrix} \mathbb{I}_2 & 0 \\ 0 & -\dfrac{\gamma_+}{\mathrm{i}}\sigma_z \end{pmatrix}, \quad \gamma_\pm = \frac{c\hbar}{W - V \pm mc^2} \tag{5.27}$$

Performing the transformation $\tilde{T}_{\Omega,0} = C_R T_{\Omega,\mathbf{k}_\parallel} C_L^{-1}$ gives

$$\tilde{T}_{\Omega,\mathbf{k}_\parallel} = \begin{pmatrix} \cos k_z d\,\mathbb{I}_2 & -\dfrac{\sin k_z d}{k_z}\,\mathbb{I}_2 \\ -k_z \sin k_z d\,\mathbb{I}_2 & -\cos k_z d\,\mathbb{I}_2 \end{pmatrix}, \qquad \mathbf{k}_\parallel = 0 \tag{5.28}$$

where we have used the fact that

$$\frac{1}{\gamma_-\gamma_+} = \frac{(W - V - mc^2)(W - V + mc^2)}{c^2\hbar^2} = \frac{(W-V)^2 - (mc^2)^2}{c^2\hbar^2} = k_z^2. \tag{5.29}$$

The transfer matrix, $\tilde{T}_{\Omega,0}$ is diagonal in spin and is identical to the non-relativistic transfer matrix, $t_{\Omega,0}$ for the same system, given by [118]

$$t_{\Omega,0} = \begin{pmatrix} \cos k_z d & -\dfrac{\sin k_z d}{k_z} \\ -k_z \sin k_z d & -\cos k_z d \end{pmatrix}, \tag{5.30}$$

and we see that the relativistic transfer matrix has the correct non-relativistic limit when the momentum is along the spin-quantisation axis.

A further useful property of the transfer matrix is illustrated by considering, for simplicity, the free space relativistic transfer matrix in 1-dimension, $T_{\Omega,0}$, given by

$$T_{\Omega,0} = \begin{pmatrix} \cos k_z d\,\mathbb{I}_2 & \dfrac{\mathrm{i}}{\gamma_+ k_z}\sigma_z \sin k_z d \\ \dfrac{\mathrm{i}}{\gamma_- k_z}\sigma_z \sin k_z d & \cos k_z d\,\mathbb{I}_2 \end{pmatrix} \tag{5.31}$$

where we continue to use $\mathbf{k}_\parallel = 0$ for transparency. The eigenvalues, $\lambda_{\pm,m_s}$ of $T_{\Omega,0}$ are

$$\lambda_{\pm,m_s} = \cos k_z d \pm \mathrm{i} \sin kd = \mathrm{e}^{\pm \mathrm{i} k_z d} \tag{5.32}$$

with (unnormalized) eigenvectors

$$U_{\pm,m_s} = \begin{pmatrix} \phi_{m_s} \\ \pm\gamma_+ \sigma_z k_z \phi_{m_s} \end{pmatrix} \tag{5.33}$$

The eigenvectors are the surface values of a relativistic plane wave on the surface at $z = 0$ and the eigenvalues are the twofold spin-degenerate pair $\mathrm{e}^{\pm \mathrm{i} k_z d}$. We will

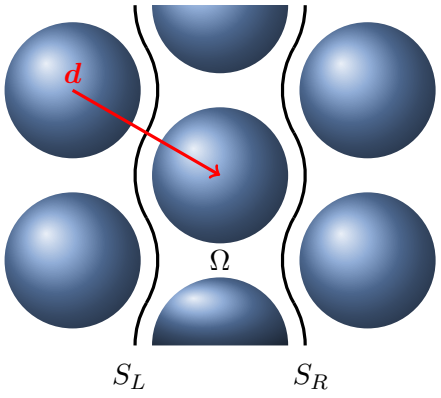

Figure 5-6: Geometry for calculating the transfer matrix for a bulk crystal. Adjacent layers are related by a translation **d** which takes surface S_L through region Ω to S_R.

expand on this idea in the next section.

5.3 The Transfer Matrix and Complex Band Structures

Having investigated the general transfer matrix, we will now concentrate on a specific case in which the region, Ω, is a representative layer of a bulk crystal with three-dimensional translational symmetry as shown in figure 5-6. This layer can reproduce the bulk crystal by repetitions normal to the layers (the z-direction) with an associated layer-layer translation vector, **d**, which reflects the bulk translational symmetry and takes a point on surface S_L to an equivalent point on S_R.

Bloch states in the bulk crystal satisfy

$$\Psi(\mathbf{r}_R) = \Psi(\mathbf{r}_L + \mathbf{d}) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{d}}\Psi(\mathbf{r}_L) \tag{5.34}$$

where $\mathbf{r}_R$ is on S_R and $\mathbf{r}_L$ is on S_L. The transfer matrix for region Ω relates the boundary values of the wavefunction on S_L and S_R

$$\Psi(\mathbf{r}_R) = \int_{S_L} \mathrm{d}^2\mathbf{r}_L T_\Omega(\mathbf{r}_R, \mathbf{r}_L)\Psi(\mathbf{r}_L) \tag{5.35}$$

and comparing with (5.34) gives the eigenvalue problem

$$\int_{S_L} \mathrm{d}^2\mathbf{r}'_L\, T_\Omega(\mathbf{r}_L + \mathbf{d}, \mathbf{r}'_L)\Psi(\mathbf{r}'_L) = \lambda\,\Psi(\mathbf{r}_L). \tag{5.36}$$

Eigenvectors of T_Ω are therefore Bloch states on S_L with eigenvalues

$$\lambda = \mathrm{e}^{\pm\mathrm{i}\mathbf{k}\cdot\mathbf{d}}. \tag{5.37}$$

This is comparable with the properties of the free particle transfer matrix in the previous section. We also note that if T_Ω transfers Ψ_L to Ψ_R, then the inverse

transfer matrix, T_Ω^{-1} maps Ψ_R onto Ψ_L.

5.3.1 Complex Bands

In a crystal with infinite translational symmetry, only eigenvectors of the translation operator, the Bloch states with real wavevectors $\mathbf{k}$ are allowed, and their variation with energy gives the band structure of the crystal. The restriction to real $\mathbf{k}$ is due to the requirement that the wavefunction should be normalizable over the crystal. If the wavevector in the Bloch phase factor $\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}$ is complex, $\mathbf{k} = \mathbf{k}_R \pm \mathrm{i}\mathbf{k}_I$ then there are wavefunctions which grow or decay exponentially through the crystal, $\mathrm{e}^{\pm\mathbf{k}_I\cdot\mathbf{r}}$, and they cannot be normalised. If the infinite translational symmetry is broken, for example by a surface, then states which decay away from the surface into the crystal can be normalised and the requirement that $\mathbf{k}$ is real no longer holds. The variation with energy of wavevectors which are in general complex is the complex band structure. The bands extend into the complex plane whenever there is a maximum or minimum in the real bands, and either connect back to another extremum in the real bands in a closed loop or extend to infinity as approximately parabolic free electron-like bands [146], [147].

The complex bands are particularly useful for investigating surface and interface states and transport properties. We have already seen that surface states correspond to electrons which are localised in the surface region. This means that their wavefunctions decay into the crystal and they must have an imaginary part in the wavevectors of the Bloch states that make up the wavefunction in the region away from the surface where the potential becomes bulk-like normal to the surface $k_z = k_z^R + \mathrm{i}k_z^I$. These surface states cannot have wavevectors in the energy range of Bloch states if they are truly localised, and surface states lie in the energy gaps of the real bands [147] [148]. From (5.34) we see that the magnitude of the imaginary part of the wavevector determines the decay of the localised surface states into the bulk crystal.

The implication of complex wavevectors for transport calculations can be illustrated by considering the elementary problem of tunneling through a one-dimensional barrier of height V and width d. In this case, the wavevector is a function of the energy, in the non-relativistic case $k = \sqrt{2(E-V)}$. If the energy of the wavefunction is above the barrier, $E > V$ and the wavevector is real, while for states with energy $E < V$, k is imaginary and the probability of tunneling through the barrier is proportional to e^{-2kd} and the wavefunction decays exponentially inside the barrier [30]. This idea can be extended to more complicated systems with the complex wavevector of the Bloch states playing the role of k in the simple model. As a consequence, complex band structures have been used, for example, to investigate magnetic tunnel junctions, [118], [149], molecular electronic devices [150], quantum transport [151] and carbon nanotubes [152].

We can easily connect the transfer matrix, T_Ω, to the complex bands by first noting that it is constructed by the Green's function at energy W for region Ω and so depends on W and we can write $T_\Omega(W)$. The eigenfunctions of $T_\Omega(W)$ are the boundary values of the Bloch states in the crystal with eigenvalue $\mathrm{e}^{\mathbf{i}\mathbf{k}\cdot\mathbf{d}}$. The real wavevectors $\mathbf{k}$ clearly give the real bands of the crystal, $W(\mathbf{k})$. For the real bands, we have already seen that time reversal symmetries of an infinite crystal lead to the property of the real bands that $W(\mathbf{k}) = W(-\mathbf{k})$ [5]. This means that the eigenvalues of T_Ω occur in pairs λ and λ' such that

$$\lambda = \mathrm{e}^{\mathbf{i}\mathbf{k}\cdot\mathbf{d}}, \qquad \lambda' = \mathrm{e}^{-\mathbf{i}\mathbf{k}\cdot\mathbf{d}}, \qquad \lambda\lambda' = 1 \tag{5.38}$$

However, λ' is an eigenvalue of the inverse transfer matrix T_Ω^{-1}, hence the eigenvalue spectrum of T_Ω and T_Ω^{-1} is the same. If we now generalise to include the complex wavevectors then we can classify the eigenvalues by $|\lambda|$. Those eigenvalues with $|\lambda| = 1$ are the Bloch states which give rise to the real bands. Those with $|\lambda| > 1$ or $|\lambda| < 1$ are states which decay within the crystal, described by wavevectors with a non-zero imaginary part in the complex band structure. We will now show how this can be applied to the bulk crystal in an embedding calculation.

5.3.2 Complex Bands from an Embedded Layer

To calculate the complex bands for a bulk crystal we need to generate the transfer matrix, T_Ω, for our representative layer. This is constructed from the Green's function with zero amplitude small component on S_L and S_R, which can be found from an embedding calculation with embedding potentials set to zero in S_L and S_R. In order to simplify the LAPW expansions of the Green's function and transfer matrix, we can transfer the boundary condition on S_L and S_R to planes P_L and P_L (see figure 5-7). The first auxiliary region Ω_1 is made by adding buffer volumes Δ_L and Δ_R to either side of Ω, transferring the boundary conditions on the curvy surfaces S_L and S_R to the planes P_L and P_R. The second auxiliary volume, Ω_2 consists of Δ_L and Δ'_R, where Δ'_R is $\Delta_R - \mathbf{d}$. This is essentially the same as the setup for the iterative method of generating the embedding potential outlined in section 5.1.2.

The transfer matrices for the auxiliary volumes can be written as products of the transfer matrices for the buffer regions and Ω

$$T_{\Omega_1} = T_{\Delta_R} T_\Omega T_{\Delta_L}, \qquad T_{\Omega_2} = T_{\Delta'_R} T_{\Delta_L}. \tag{5.39}$$

Note the ordering, since the transfer matrices act on wavefunctions to the right and transfer the boundary conditions from left to right, $P_L \to P_R$. If some general transfer matrix transfers the boundary condition from surface $S_L \to S_R$ then its inverse takes us from $S_R \to S_L$ and we can define a new transfer matrix T such that

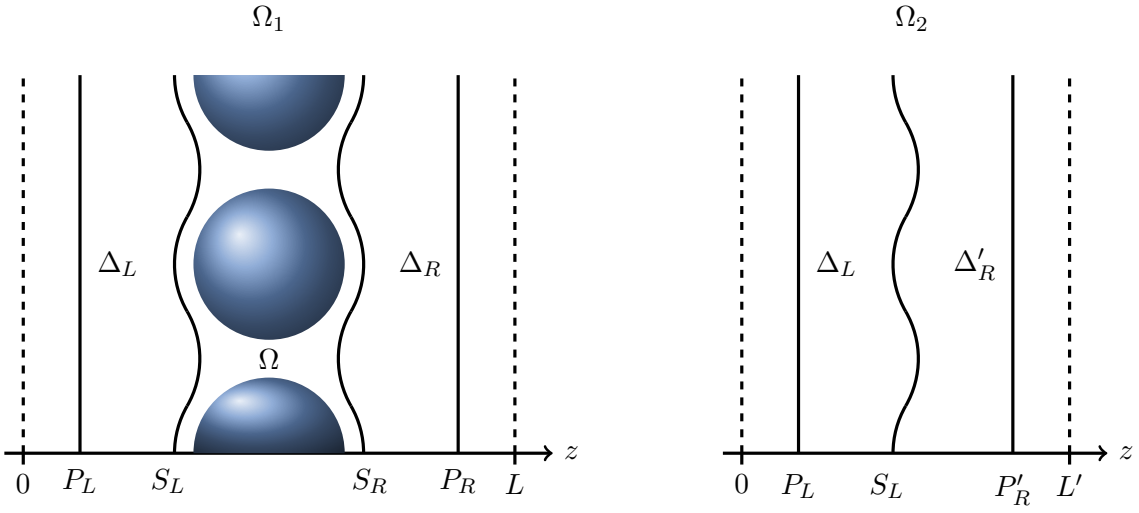

Figure 5-7: Geometry for generating the bulk transfer matrix from an LAPW embedding calculation. Volume Ω_1 consists of Ω, between S_L and S_R extended by buffer regions Δ_L and Δ_R to planes P_L and P_R. Volume Ω_2 has Δ_L and Δ'_R bounded by planes P_L and P'_R. Δ_R and Δ'_R are related by a layer-layer bulk lattice vector $\mathbf{d}$.

$$T = T_{\Omega_1} T_{\Omega_2}^{-1} = T_{\Delta_R} T_{\Omega} T_{\Delta'_R}^{-1}. \tag{5.40}$$

The first matrix, $T_{\Delta'_R}^{-1}$, takes us from P'_R to S_L, the second through Ω to S_R and the third from S_R to P_R. The eigenvectors of T_Ω are the boundary values of Bloch states on S_L, therefore the eigenvectors of T are the eigenvectors of T_Ω on S_L transferred to P'_R.

For the geometry in figure 5-7, quantities on the planes can be expanded in a set of plane wave basis functions

$$\chi_\mu = \sqrt{\frac{1}{\mathbb{A}}} \mathrm{e}^{\mathrm{i}(\mathbf{k}_\parallel + \mathbf{g})\cdot \mathbf{r}_\parallel} \phi_{m_s}, \qquad \mu = (\mathbf{g}, m_s) \tag{5.41}$$

where $\mathbb{A}$ is the area of unit cell parallel to P_L, $\mathbf{g}$ is a two-dimensional reciprocal lattice vector and the parallel momentum $\mathbf{k}_\parallel$ is in the first two-dimensional Brillouin zone. We neglect the small component of the plane waves since the Green's function used to define the transfer matrix has zero small components on the planes. The bulk complex band structure at $\mathbf{k}_\parallel$ is found by solving the eigenvalue equation

$$\sum_{\mu'} \mathrm{e}^{\mathrm{i}\mathbf{g}\cdot\mathbf{d}_{||}} T_{\mu\mu'}(W, \mathbf{k}_\parallel) U_{\mu'} = \mathrm{e}^{\mathrm{i}k_z d_z} U_\mu \tag{5.42}$$

where the phase factor $\mathrm{e}^{\mathrm{i}\mathbf{g}\cdot\mathbf{d}_{||}}$ must be included when the translation $\mathbf{d}$ is not normal to P_L. The eigenvectors, U, are the surface values of the bulk Bloch states at $\mathbf{k}_\parallel$ and their eigenvalues give the complex bands.

The eigenvalues, $\lambda = \mathrm{e}^{\pm \mathrm{i} k_z d_z}$ of T with $|\lambda| = 1$ are part of the conventional band structure, with k_z real, and describe Bloch states travelling towards $+\infty$ for $k_z > 0$ and $-\infty$ with $k_z < 0$. Eigenvalues with $|\lambda| \neq 1$ are evanescent states, k_z is complex, which decay towards $+\infty$ for $|\lambda| < 1$ and $-\infty$ if $|\lambda| > 1$.

The advantages of calculating the complex bands from the transfer matrix are that we only need to explicitly consider a single layer of bulk, we do not need to know the specific form of the curvy surfaces S_L and S_R and we have Green's functions and transfer matrices defined on planar surfaces, allowing for simple plane wave expansions.

5.3.3 Results: Some Fully Relativistic Complex Band Structures

Before presenting the results of relativistic complex band structure calculations in the LAPW basis, there are some points regarding the choice of basis functions which should be explained. Wavevectors for the LAPW basis functions are

$$\mathbf{k} = \mathbf{k}_{\|} + \mathbf{g} \pm k_n \hat{\mathbf{z}} \tag{5.43}$$

where $\mathbf{g}$ is a two-dimensional reciprocal lattice vector. The wavevectors normal to P_L, in the z-direction, are defined over the distances L and L' in order to retain variational freedom in the basis;

$$k_n = \frac{2n\pi}{L} \quad \text{in} \quad \Omega', \qquad k_n = \frac{2n\pi}{L'} \quad \text{in} \quad \Omega''. \tag{5.44}$$

The sizes of the two volumes in the z-direction are related by $P_R - P_L = P'_R - P'_L + d_z$. The potential in the buffer regions is arbitrary but for practical purposes can be chosen to be a smooth extension of the interstitial potential.

In a typical LAPW surface calculation, the number of plane wave basis functions is determined by defining a maximum 'cut-off' 3-dimensional reciprocal lattice vector $\mathbf{G}_{max}$ and the $\mathbf{k}$ are chosen to lie within a sphere of radius $\mathbf{G}_{max}$ for a given $\mathbf{k}_{\|}$ [118]

$$\{|\mathbf{g} + k_n \hat{\mathbf{z}}|\} \leq |\mathbf{G}_{max}|. \tag{5.45}$$

This means that the number of k_n, or n_{max}, is determined by $\mathbf{G}_{max}$ and the number of k_n for a given $|\mathbf{g}|$ decreases as we approach the cutoff radius $\mathbf{G}_{max}$. The transfer matrix used for the bulk complex band structure is constructed from projections of the Green's function for region Ω onto the embedding planes P_L and P_R. The complex band structure is therefore dependent on the accuracy of these surface projections. The lack of a sufficient number of k_n close to $\mathbf{G}_{max}$ can lead to an insufficient representation of the transfer matrix, resulting in unphysical Bloch states in the band structure due to incorrect eigenvectors of the transfer matrix. This is because strongly decaying evanescent states with very large two dimensional $\mathbf{g}$

vectors are inadequately described and appear as Bloch states in the real band structure [118]. To remedy this problem, the two-dimensional $\mathbf{g}$ are chosen from a circle of radius $\mathbf{G}_{max}$ and the k_n are chosen over the length L in figure 5-7, up to some chosen n_{max}

$$\left\{ |\mathbf{g}| \leq |\mathbf{G}_{max}|_{||}, k_n = \frac{2\pi n}{L}, n \leq n_{max} \right\} \tag{5.46}$$

leading to a cylindrical basis set. This basis contains larger k_n than the spherical basis and some care has to be taken when constructing the matrix elements because large $\mathbf{g}$ correspond to high kinetic energies and these contributions must also be included in the Hamiltonian in the atomic spheres. One way to remedy this is to use larger angular momentum cutoffs inside the spheres [118]. The approach used in the calculations presented in this work is that of Soler and Williams [153]. Instead of increasing the angular momentum cutoff in the spheres, the plane wave matrix elements are calculated throughout the whole region (interstitial and atomic spheres) and the contribution from the plane waves is then subtracted from the atomic spheres. This is achieved by expanding the plane waves in spherical waves up to the same angular momentum cutoff as the atomic solutions. The contribution to the high $\mathbf{g}$ vectors is accounted for in the LAPW expansion coefficients.

First we calculate the Hamiltonian and overlap matrix elements for volumes Ω_1 and Ω_2 for a chosen parallel wavevector $\mathbf{k}_{\|}$, in the first Brillouin zone. Next, the Green's function for each volume is calculated over a range of energies, W

$$[G_{\Omega_i}(W)]_{\nu\nu'} = [WO_{\Omega_i} - H_{\Omega_i}]^{-1}_{\nu\nu'}, \quad \nu = (\mathbf{k}, m_s) \tag{5.47}$$

where W is typically chosen to be complex with a small imaginary part. To construct the transfer matrix, we use

$$T_{\Omega_i} = \begin{pmatrix} F_{RR}F_{LR}^{-1} & F_{RL} - F_{RR}F_{LR}^{-1}F_{LL} \\ \\ F_{LR}^{-1} & -F_{LR}^{-1}F_{LL} \end{pmatrix} \tag{5.48}$$

where $F_{LL} = -\mathrm{i}c\hbar G_{ll}(\mathbf{r}_L, \mathbf{r}'_L)\sigma_L$ etc. We need to expand the large-large components of the Green's functions on the planes P_L etc. To do this we use the large component of the plane wave LAPW basis functions

$$\begin{aligned} \left[G_{\Omega_i}(Z_S, Z_{S'}; W, \mathbf{k}_{\|})\right]_{\mu\mu'} &= \frac{2}{\mathbb{V}} \sum_{nn'} \int_{\mathbb{A}} \mathrm{d}^2\mathbf{r}_{||} \psi_\nu(\mathbf{r}) \left[G_{\Omega_i}(W)\right]_{\nu\nu'} \psi^{\times}_{\nu'}(\mathbf{r}) \\ \\ &= \frac{2}{\mathbb{L}} \sum_{nn'} \delta(\mathbf{g} - \mathbf{g}') \left[G_{\Omega_i}\right]_{\nu\nu'} \sin(k_n Z_S) \sin(k_{n'} Z_{S'}). \end{aligned} \tag{5.49}$$

The sum is over k_n and $k_{n'}$ and the index $\mu = (\mathbf{g}, m_s)$ labels only the reciprocal lattice vectors and the spin. The sin dependence is replaced with cosines for cos-like basis functions and Z_S and Z'_S are the z positions of the planes P_L, P_R and so on for the F_{LL} etc. The distance $\mathbb{L}$ is the distance between the planes at in Ω_i, not the distance L over which the k_n are defined. The σ_S matrices in the $F_{SS'}$ are just σ_z and we must be careful to include the correct sign for the normal derivative, -1 for Z'_S on P_L and $+1$ for Z'_S on P_R. Substituting into (5.48) gives the transfer matrix at energy W for volume Ω_i indexed by reciprocal lattice vectors and spins

$$\begin{bmatrix} T^{ll}_{\Omega_i}(W) & T^{ls}_{\Omega_i}(W) \\ \\ T^{sl}_{\Omega_i}(W) & T^{ss}_{\Omega_i}(W) \end{bmatrix}_{\mu\mu'}, \qquad \mu = (\mathbf{g}, m_s). \tag{5.50}$$

We then use (5.40) to construct the transfer matrix T for Ω and solve (5.42) for the eigenvalues to get the complex bands.

Figures 5-9 - 5-12 show scalar relativistic and fully relativistic complex band structures for fcc gold along the Γ-X (001) and Γ-L (111) directions. Figure 5-8 shows the fcc Brillouin zone, with symmetry labels (reproduced from [154]).

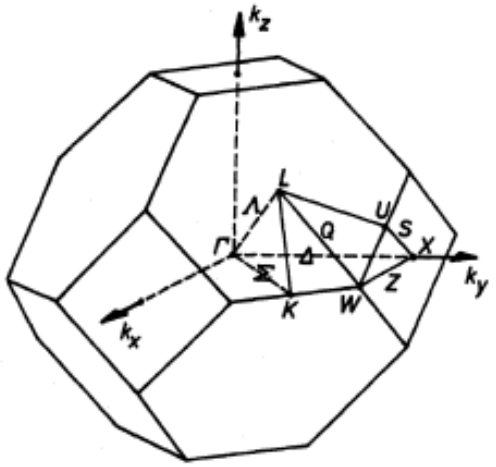

Figure 5-8: The Brillouin zone for an fcc lattice, showing symmetry points.

In each of the figures, the red lines are the real bands, with $\mathrm{Im}(k) = 0$ and the central panel shows k increasing from left to right, from the Γ point, at the centre of the Brillouin zone, along Δ to X at the centre of the square (001) face, or along Λ to L at the centre of the hexagonal (111) face (see figure 5-8). The Green lines in the left panels show the purely imaginary bands with $\mathrm{Re}(k) = 0$. Only those with $\pi \geq \mathrm{Im}(k) \geq 0$ are shown for clarity, but are identical to those with $-\pi \leq \mathrm{Im}(k) =\leq 0$. The Green lines in the right panels are the bands with $\mathrm{Re}(k)$ at either X or L, and again are symmetrical about the real axis but are shown only for

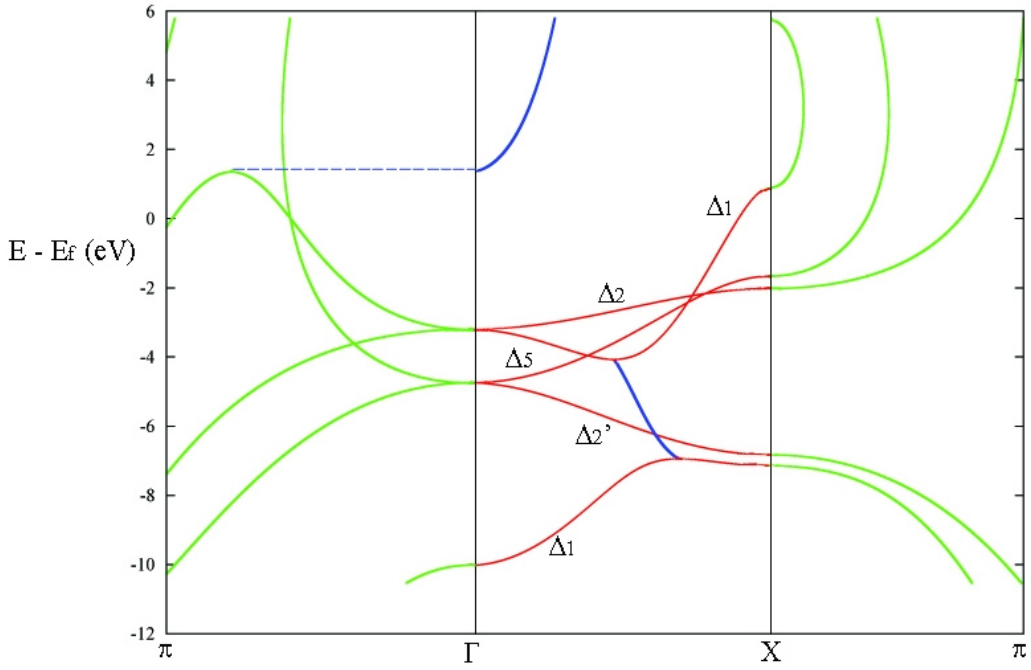

Figure 5-9: The scalar relativistic complex band structure of Au along Γ-X (001).

$0 \geq \mathrm{Im}(k) \geq 0$. The blue lines in the central panel are the general complex bands projected on to the real axis.

The real bands are labelled by their symmetry and these labels show how the Bloch states for each band transform under the symmetry operations of the point group of the Brillouin zone along a particular direction. For example, at the zone centre, the Γ point, the Brillouin zone has full octahedral symmetry, O_h, but as we move from Γ to X, along Δ, the group changes to the point group of a square, C_{4v}, and then to the group D_{4h} (4-fold rotations and mirror planes) at X [5]. Wavefunctions must transform as irreducible representations of the particular symmetries of the Brillouin zone [132] and so at the centre, we can take linear combinations of the s, p and d orbitals which generate the irreducible representations of O_h. Along Δ, these linear combinations will not, in general have the appropriate symmetries and we will have to take new ones to generate the irreducible representations of C_{4v}. This means that the degeneracies due to the linear combinations at Γ will be lifted and the bands will split. For example, at Γ the three p-orbitals are symmetrically equivalent, but as we move along the k_z axis, combinations of p_x and p_y transform as a different representation of C_{4v} to p_z. As a result, the Bloch states which give rise to the energy bands will have different symmetries along different crystal directions, and the bands can be labelled by these symmetries.

The only significant difference in the scalar-relativistic versus fully relativistic calculations is the inclusion of spin-orbit interaction in the relativistic calculation.

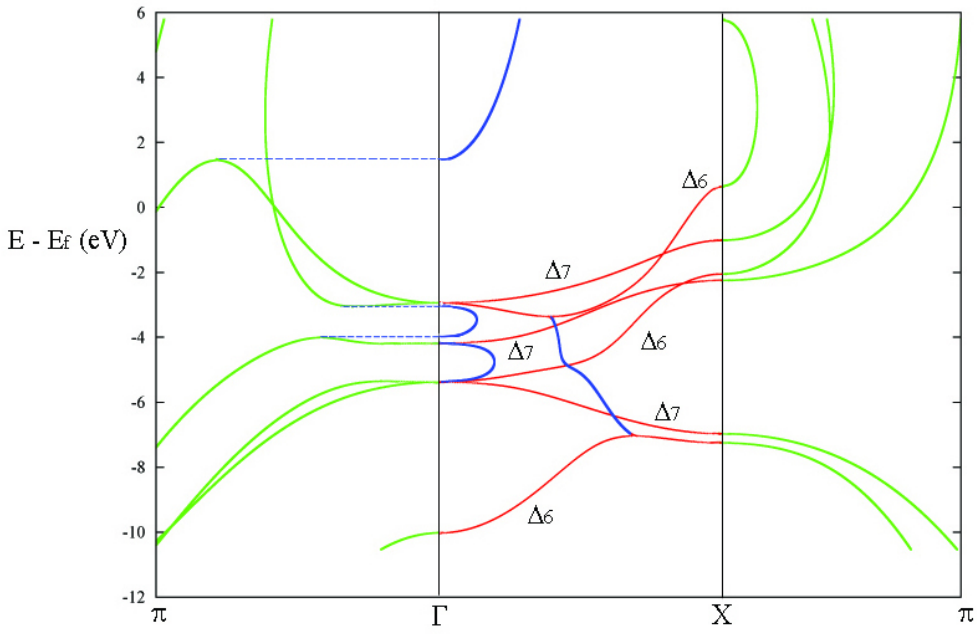

Figure 5-10: Fully relativistic complex band structure of Au along Γ-X (001).

The effect of this is clearly visible when comparing the complex bands. In the relativistic band structures, we see that the spin-orbit interaction lifts the degeneracies of some of the bands due to the different symmetries of the wavefunctions, but does not produce a spin-splitting because of the inversion symmetry present in the bulk crystal.

General features of the complex bands are the presence of free electron like, almost parabolic bands continuing from the real bands into the complex plane (Green lines). It is clear that there are also bands in these regions which are closed loops connecting real bands with the same symmetry across the band gaps near the Fermi level at X and L. There are also complex bands connecting extrema of the real bands. For example, in the relativistic bands along Au(111), a line connects the minimum of the Λ_4 band near -3 eV to the maximum of the Λ_4 band at -4 eV. In the non-relativistic bands along Au(001) we see that there is also a loop in the complex plane connecting, at $\approx$ 1.5 eV, to a maximum of the purely imaginary line which continues from the joining of the Δ_2 and Δ_5 at the Γ point, at around -3 eV. A similar feature appears in the relativistic bands along Γ-X, along with a second loop connecting a maximum and minimum of the imaginary bands continuing the Δ_7 and Δ_6 bands near -3 eV, and the Δ_7 band at Γ near -4 eV.

The calculations were performed with the cylindrical basis consisting of 673 2-d reciprocal lattice vectors, up to a cut-off length of 18.1 r_B^{-1} and 30 k_n. The lattice parameter was 4.08 Å, and the distance, L, over which the basis functions are defined was 8.7 r_B (see figure 5-7). Embedding planes were placed at $P_L = 1.5\, r_B$ and

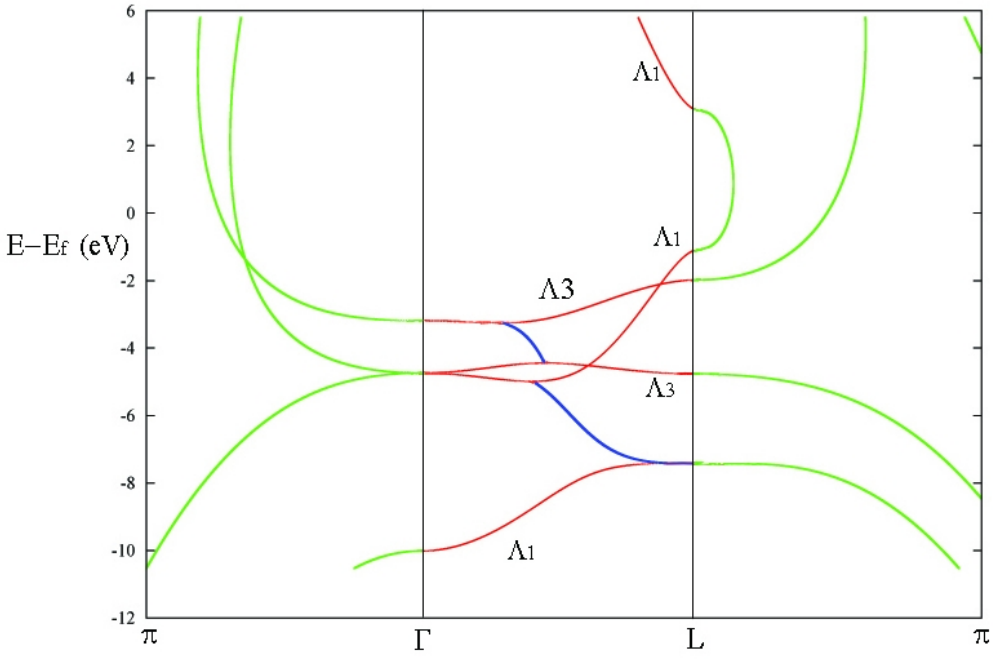

Figure 5-11: Scalar relativistic complex band structure of Au along Γ-L (111).

$P_R = 7.2\ r_B$, chosen so that they gave sufficient interstitial distance on either side of the atomic sphere, radius 2.72 r_B, centred at z=4.35. 8016 energy points were used over the range -0.1 to 0.5 Hartrees and a small imaginary part 1.5×10^{-4} Hartrees was added to each energy. The bulk potential and Fermi energy came from a scalar-relativistic APW calculation in the LDA approximation using the Perdew-Zunger parameterization of the exchange-correlation potential.

5.4 The Transfer Matrix and the Bulk Embedding Potential

At the beginning of this chapter we discussed some ways of generating the bulk embedding potential. The first method was in terms of the reflection matrix for a bulk LKKR calculation, but there are problems with using this for a full potential LAPW surface calculation due to the muffin tin approximation. We also discussed the iterative methods, either using an embedding plane located half-way between atomic layers or by using two auxiliary volumes consisting of buffer regions similar to those used for calculating the complex bands via the transfer matrix. Iterative methods for generating the embedding potential are not ideal since they are sensitive to the input 'guess' for the embedding potential, which is often generated from a separate calculation [105], or can converge poorly over certain energy ranges in the band structure or for energies with a very small imaginary part [105], [118]. In this section we will describe how the bulk embedding potential that replaces a bulk

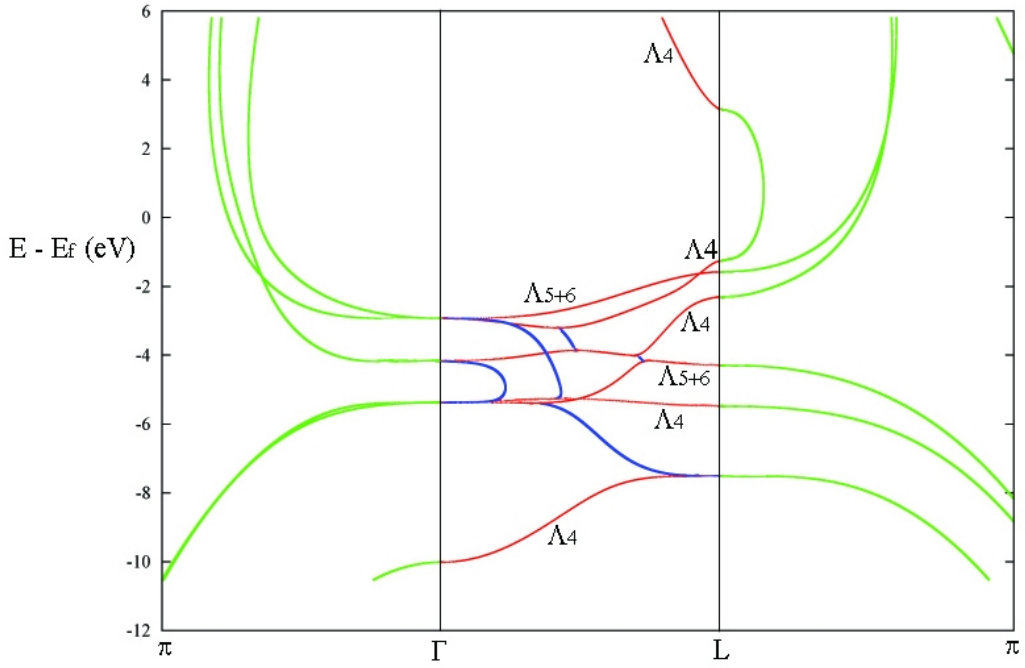

Figure 5-12: Fully relativistic band structure of Au along Γ-L (111)

substrate may also be calculated from the transfer matrix.

Recall that the embedding potential at energy W relates the amplitude of large and small components of the wavefunction on the embedding surface S

$$\psi_s(\mathbf{r}_S) = \mathrm{i}c\hbar \int_S \mathrm{d}^2\mathbf{r}'_S \cdot \Gamma(\mathbf{r}_S, \mathbf{r}'_S; W)\boldsymbol{\sigma}\psi_l(\mathbf{r}'_S). \tag{5.51}$$

For our embedded bulk layer in figure 5-7, having introduced the buffer region Δ_L, the embedding potential for the crystal to the left of S_L is transferred to the plane P_L where it can be expanded in plane waves over P_L

$$\Gamma(\mathbf{r}_L, \mathbf{r}'_L; W, \mathbf{k}_\parallel) = \frac{1}{\mathbb{A}} \sum_{\mu\mu'} \left[\Gamma(W, \mathbf{k}_\parallel)\right]_{\mu\mu'} \chi_\mu(\mathbf{r}_L)\chi^\dagger_{\mu'}(\mathbf{r}'_L). \tag{5.52}$$

where we have used the basis functions in (5.41). This matrix representation of Γ gives a set of linear equations which allow Γ to be determined as follows.

For the bulk crystal, the embedding potential can be used to relate the boundary values of the large and small components of Bloch states incident on P_L. The eigenvectors of the transfer matrix T are the boundary values of Bloch states on P'_R. To construct the embedding potential on P_L to give the correct outgoing boundary condition we need the Bloch states on P_L which decay and travel away from the surface, i.e. those with eigenvalue $|\lambda| \geq 1$ and $k_z < 0$. The inverse transfer matrix for volume Ω_2 transfers the boundary condition from P'_R to P_L, so if U_j is an eigenvector of T on P'_R then $\psi_j = T^{-1}_{\Omega_2} U_j$ is a Bloch state on P'_R back projected through Ω_2 onto

P_L.

If we write the eigenvectors, ψ_j on P_L as the columns of matrices consisting of large and small components of ψ_j

$$\left[\Psi^l\right] = \left(\psi_1^l, \psi_2^l, \cdots\right), [\Psi^s] = \left(\psi_1^s, \psi_2^s, \cdots\right), \tag{5.53}$$

then the equation relating large and small components via the embedding potential becomes, in symbolic matrix notation,

$$[\Psi^s] = -\mathrm{i}c\hbar\left[\Gamma(W)\right]\sigma_z\left[\Psi^l\right]. \tag{5.54}$$

Since the eigenvectors form a linearly independent set, we can rearrange this equation to solve for the embedding potential

$$[\Gamma(W)] = \frac{\mathrm{i}}{c\hbar}\left[\Psi^s\right]\left[\Psi^l\right]^{-1}\sigma_z. \tag{5.55}$$

We must of course choose the eigenvectors which correspond to the outgoing boundary condition on P_L; those which decay or travel away from the surface into the crystal. These can be sorted by their eigenvalues as discussed in relation to the complex bands.

5.4.1 The Bulk Embedding Potential in the LAPW Basis

So far we have described the method for generating the embedding potential from the transfer matrix in a rather general way. We will now be more specific and show how it can be achieved with a relativistic LAPW basis set.

First, we calculate the eigenvalues, λ_j and eigenvectors $U_{\mu j}$ of $\left[T(W, \mathbf{k}_\parallel)\right]_{\mu\mu'}$ (see equation 5.42) on P_R'. There are $4N$ eigenvalues, corresponding to a pair $\mathrm{e}^{\pm\mathrm{i}k_z d_z}$ for each spin m_s. If we have a basis consisting of N g-vectors, then each eigenvector $U_{\mu j}$, $\mu = (\mathbf{g}, m_s)$, will be a column vector with $4N$ rows; a large and small component for each of the two spins m_s. The eigenvectors are transferred back through Ω_2 onto P_L,

$$\sum_{\mu'}\left[T_{\Omega_2}^{-1}\right]_{\mu\mu'}[U_j]_{\mu'} = [\psi_j]_\mu \tag{5.56}$$

to give the boundary values of the Bloch states on P_L, $\psi_{\mu j}$.

We now choose the states satisfying the outgoing boundary condition on P_L by examining the eigenvalues. It is straightforward to separate out those states which decay away from P_L, as these have eigenvalues with $|\lambda_j| > 1$. More difficult are the travelling states, since states corresponding to the real bands have $|\lambda_j| = 1$, which does not determine whether they travel away from or towards P_L. For these we calculate their group velocity normal to P_L, v_j, as the states we want have negative

v_j. The relativistic velocity operator is $c\boldsymbol{\alpha}$ and we calculate v_j from

$$v_j = \sum_{\mu\mu'} [\psi_j]^\dagger_\mu [\alpha_z]_{\mu\mu'} [\psi_j]_{\mu'} , \tag{5.57}$$

where

$$[\alpha_z] = \begin{pmatrix} 0 & [\sigma_z]_{\mu\mu'} \\ [\sigma_z]_{\mu\mu'} & 0 \end{pmatrix}, \qquad [\sigma_z]_{\mu\mu'} = \delta_{\mathbf{g}\mathbf{g}'}\delta_{m_s m'_s}\left(\delta_{m_s\,\frac{1}{2}} - \delta_{m_s\,-\frac{1}{2}}\right). \tag{5.58}$$

In this way, the $4N$ states are reduced to $2N$ outgoing states travelling or decaying away from P_L into the crystal. The Bloch states satisfying the appropriate outgoing boundary condition are then put into the $2N \times 2N$ matrices of large and small components $\left[\Psi^l\right]_{\mu j}$ and $[\Psi^s]_{\mu j}$ and ,finally, we construct the matrix representation of the embedding potential on P_L through

$$\left[\Gamma(W, \mathbf{k}_\parallel)\right]_{\mu\mu'} = \frac{\mathrm{i}}{c\hbar} \sum_{j\mu''} [\Psi^s]_{\mu j} \left[\Psi^l\right]^{-1}_{j\mu''} [\sigma_z]_{\mu''\mu'} . \tag{5.59}$$

This is the embedding potential for the semi-infinite bulk crystal to the left of S_L transferred through the buffer region Δ_L on to P_L and contains all the information about the Bloch and evanescent waves satisfying the outgoing boundary condition on S_L (i.e. the complex band structure). If we use this embedding potential to represent the substrate in an embedded surface calculation, we must include the buffer region Δ_L in the embedded surface region to ensure that we impose the correct boundary condition on the 'real' embedding surface S_L.

5.4.2 The Vacuum Embedding Potential

At this point, having discussed the construction of the embedding potential describing a semi-infinite substrate from the transfer matrix, it is convenient to extend the discussion to the other embedding potential needed in an embedded surface calculation. It is assumed that the vacuum level, V_0 is constant and we can derive an analytic expression for the embedding potential on the plane P. To do so, we note that wavefunctions in the constant potential vacuum are relativistic plane waves

$$\psi(\mathbf{r}) = \begin{pmatrix} \phi_{m_s} \\ \gamma\boldsymbol{\sigma}\cdot\mathbf{k}\phi_{m_s} \end{pmatrix} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}, \tag{5.60}$$

where

$$\mathbf{k} = \mathbf{k}_\parallel + \mathbf{g} + k_z\hat{\mathbf{z}}, \quad \gamma = \frac{c\hbar}{W - V_0 + mc^2}, \qquad k_z = \frac{1}{c\hbar}\sqrt{(W-V_0)^2 - (mc^2)^2 - c^2\hbar^2|\mathbf{k}_\parallel + \mathbf{g}|^2}. \tag{5.61}$$

and assume that the vacuum includes $+\infty$ so that wavefunctions satisfying the outgoing boundary condition on P correspond to $k_z \geq 0$. The embedding potential relates large and small components of the wavefunctions on the plane P and can be expanded on P as

$$\Gamma(\mathbf{r}_P, \mathbf{r}'_P; W) = \frac{1}{(2\pi)^2} \int_{\mathcal{A}} \mathrm{d}^2\mathbf{k}_\parallel \, \Gamma(W, \mathbf{k}_\parallel) \mathrm{e}^{\mathrm{i}\mathbf{k}_\parallel \cdot (\mathbf{r}_P - \mathbf{r}'_P)} \tag{5.62}$$

and we can determine $\Gamma(W, \mathbf{k}_\parallel)$ by substituting (5.60) into (5.51) to get

$$\Gamma(W, \mathbf{k}_\parallel) = \frac{-\mathrm{i}}{W - V_0 + mc^2} \left[\boldsymbol{\sigma} \cdot \mathbf{k}_\parallel \sigma_z + k_z \right]. \tag{5.63}$$

It is clear that this expression is independent of the position of the embedding plane for a region of constant potential as we would expect.

Writing out the $\boldsymbol{\sigma}$ matrices explicitly we have

$$\Gamma(W, \mathbf{k}_\parallel) = \frac{-\mathrm{i}}{W - V_0 + mc^2} \begin{pmatrix} k_z & -k_x - g_x + \mathrm{i}k_y + g_y \\ k_x + g_x + \mathrm{i}k_y + g_y & k_z \end{pmatrix}. \tag{5.64}$$

It is obvious that this vanishes as $c \to \infty$, but from section 4.3.1 we know that we must include a factor of $c^2\hbar^2$ to take the non-relativistic limit. Doing this, we get

$$\Gamma(W, \mathbf{k}_\parallel) = \frac{-\mathrm{i}\hbar^2}{2m} \begin{pmatrix} k_z & -k_x - g_x + \mathrm{i}k_y + g_y \\ k_x + g_x + \mathrm{i}k_y + g_y & k_z \end{pmatrix}. \tag{5.65}$$

The diagonal terms are identical to the non-relativistic embedding potential for free space. Once again we have the extra off diagonal terms due to the coupling of spin and momentum that we encountered in section 4.3.1.

5.5 The Embedding Potential and Ballistic Transport

Before moving on to the application of the relativistic embedding method to surface calculations, we will present a short discussion of its role in transport calculations. Recent interest in spintronics and molecular electronics means that much attention is focussed on the transport properties of thin interface layers and organic molecules or carbon nanotubes attached to semi-infinite leads.

Various levels of complexity are used to approximate electron transport in real devices, ranging from those which attempt to describe all the scattering of the electrons within the material to those which include only the most basic, dominant scattering processes.

In general, we could calculate the transition probability from one multi electron state to another as a result of the applied field but this would require understanding of the time-dependence of the many electron wavefunctions and is extremely difficult

to solve. By considering the current as a linear response to the applied field, the transport properties, including both elastic and inelastic scattering, can be calculated with the Kubo method [155], [156], which is typically formulated in terms of non-equilibrium Green's functions and is well suited to systems where impurity effects are important or interdiffusion causes surface 'roughness' [157]. This approach has been applied to a range of systems, most notably, in the context of this work, to giant magnetoresistance in multi layered systems in the KKR-CPA approximation [158], [159], [157], [160]. The disadvantage of using the Kubo formula is that it involves evaluation of time-averaged quantities and certain approximations must be made [161].

If processes such as electron-electron and electron-phonon scattering happen on average on a length scale λ, then in devices whose size is less than λ these processes will not in general occur during transport through the device. We may therefore describe the conductance in terms of reflection and transmission of single particle wavefunctions at the interface between the device and the leads and the electrons do not scatter (change their energy) in travelling across the device, therefore the wavevector, $\mathbf{k}$, is a good quantum number, this is *ballistic transport*.

5.5.1 The Landauer-Büttiker Formula

An important method of calculating conductances in terms of interface reflection and transmission coefficients is the Landauer-Büttiker formulation [162], [163], [164]. This has been applied to nanostructures in wave-function matching calculations [165], [166] where reflection and transmission coefficients are calculated by direct matching of Bloch states across the device region, and within the non-relativistic embedding framework [46], [118]. The Kubo and Landauer-Büttiker formulas are, naturally, equivalent in the appropriate limits [164].

To derive an expression for the conductance, we assume that our device, Ω, is connected on either side to reservoirs L and R and the system is in thermal equilibrium. In moving from L to R, electrons are only scattered at the interfaces between Ω and L and Ω and R and travel ballistically across Ω. A potential difference V applied across Ω results in a difference $\delta E = eV$ between the chemical potentials in L, μ_L and R, μ_R as illustrated in figure 5-13. If we consider, for the sake of simplicity, a situation in which there is a single conduction channel[1] then at low temperatures, where we can assume that states are filled up to μ_L and μ_R, the current from L to R is

$$I_{L\to R} = e\int_0^{k_F} \mathrm{d}k\, v(k), \tag{5.66}$$

[1]This could be a reasonable model of a 1-d quantum wire connected to point contact leads.

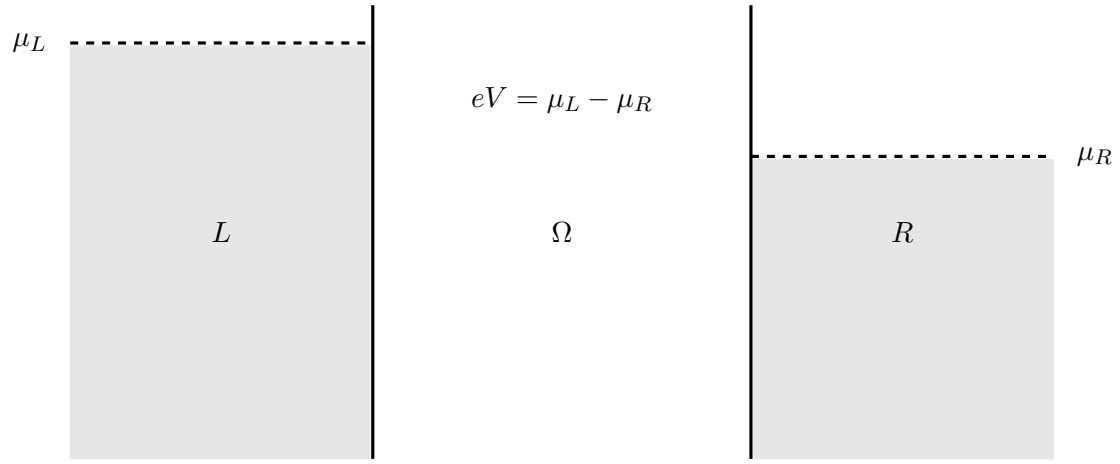

Figure 5-13: Schematic view of a device, Ω, connected to left, L, and right hand, R, leads. Electron states in the leads are filled up to the chemical potentials μ_L and μ_R. A potential difference V applied across the device causes an energy offset, δE in the chemical potentials $\delta E = eV = \mu_L - \mu_R$.

where the group velocity of the state at k is

$$v(k) = \frac{1}{\hbar}\frac{\partial E}{\partial k}. \tag{5.67}$$

The density of states is given by

$$n(E) = \frac{\partial k}{\partial E} \tag{5.68}$$

and the integral over k can be turned into an energy integral for the current from L to R

$$I_{L\to R} = \frac{2e}{\hbar}\int_0^{\mu_L} \mathrm{d}E\, \frac{\partial E}{\partial k}\frac{\partial k}{\partial E} = \frac{e}{\hbar}\mu_L, \tag{5.69}$$

where the factor of 2 comes from the spin-degeneracy of the channel.

A similar argument follows for the current $I_{R\to L}$ flowing from R to L and the conductance, C, is related to the net current,

$$C = \frac{I_{L\to R} - I_{R\to L}}{V} = \frac{2e^2}{\hbar}. \tag{5.70}$$

Thus, for a single channel, the conductance across a device where scattering processes may be ignored is a fixed value, the *conductance quantum*, equal to $7.748 \times 10^{-5}\Omega^{-1}$ (which is an equivalent resistance of $\approx 12907\Omega$).

In a system where several independent bands or conduction channels exist, the situation becomes more complex. For example, in a device connected between semi infinite leads we must consider the transmission of Bloch states in the leads across the device. If we define a coefficient, t_{ij}, describing the transmission of Bloch state i

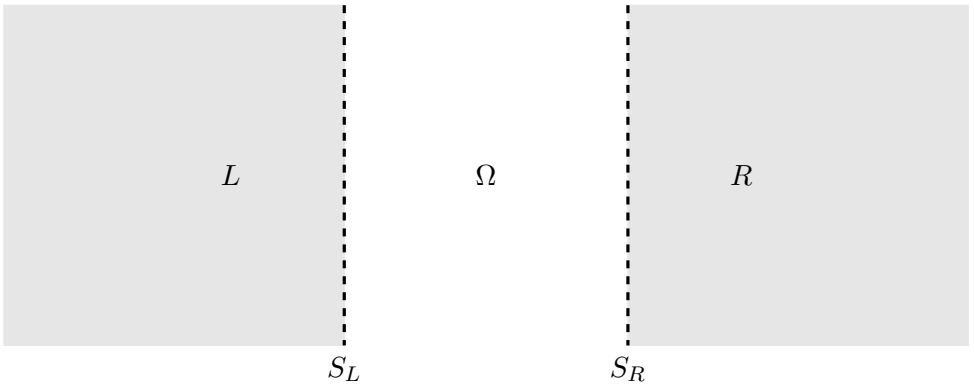

Figure 5-14: The geometry for a typical transport calculation. The 'device region', Ω is connected to semi-infinite leads L and R at the surfaces S_L and S_R.

in L, across the device into Bloch state j in R, then the conductance is given by [155]

$$C = \frac{2e^2}{\hbar} \sum_{ij} |t_{ij}|^2 , \tag{5.71}$$

where the sum is over all current carrying channels. This is the so called *two-terminal* Landauer equation for a system consisting of multiple conduction channels and describes the conductance in a situation where the potential difference and current across the device are measured at the same point in the leads L and R. In many measurements of conductance through mesoscopic devices, the potential difference and current are measured at different points and the extension of the Landauer approach to these four-terminal measurements was introduced by Büttiker [163]. However, there are many situations in which two-terminal measurements are performed [167], particularly where the devices are exceptionally small, for example quantum dots [168].

The dependence of the transmission on the available conduction channels for Bloch states means that the embedding method is particularly well suited to ballistic transport problems. Because it represents a true semi-infinite lead, the embedding potential contains all the information about these conduction channels via the complex band structure. The non-relativistic embedding method has been applied to the conductance of long chain organic molecules, namely DNA, within a tight binding framework [119] and a carbon nanotube connected to semi-infinite leads [120] . An FLAPW implementation based on the transfer matrix has also been used to study ballistic transport through vacuum and ferromagnetic tunnel junctions [46], [169], [170] and oxygen doped Au and Ag wires [171].

Many transport studies make use of a well known formula for the transmission, T, across a device region Ω, connected to semi-infinite leads L and R as in figure 5.72 [152], [46], [119], [151], [117]:

$$T = 4\,\mathrm{Tr}\,[\mathrm{Im}\Sigma_L G_{LR} \mathrm{Im}\Sigma_R G^*_{RL}]\,. \tag{5.72}$$

The Green's functions G_{LR} and G_{RL} are the Green's functions of the device region Ω with $\mathbf{r}$ and $\mathbf{r}'$ on surfaces S_L and S_R connecting the leads L and R to Ω. The quantities $\mathrm{Im}\Sigma$ are variously interpreted as the 'self-energy' on S_L and S_R or the imaginary part of the embedding potentials for the semi-infinite leads on S_L and S_R. The trace contains a sum over the conduction channels and a particular advantage of the expression (5.72) is that it is independent of the explicit representation of these channels [119]. In the remainder of this section we will show how the relativistic embedding potential leads to a natural definition of a set of surface functions suitable for defining the channels and use these to derive a relativistic equivalent of (5.72).

5.5.2 Embedding and the Surface Current

To see how the embedding potential arises naturally in a description of transport across the interface between the device region and the leads we consider the relativistic current operator, $c\boldsymbol{\alpha}$. The current across a surface S is given by

$$J_S = c\int_S \mathrm{d}^2\mathbf{r}_S \cdot \psi^\dagger(\mathbf{r}_S)\,\boldsymbol{\alpha}\,\psi(\mathbf{r}_S) \tag{5.73}$$

and we can write this in terms of the large and small components of the wavefunction and the Pauli $\boldsymbol{\sigma}$ matrices

$$J_S = c\int_S \mathrm{d}^2\mathbf{r}_S \cdot \left[\psi_l^\dagger(\mathbf{r}_S)\,\boldsymbol{\sigma}\,\psi_s(\mathbf{r}_S) + \psi_s^\dagger(\mathbf{r}_S)\,\boldsymbol{\sigma}\,\psi_l(\mathbf{r}_S)\right]. \tag{5.74}$$

The embedding potential $\Gamma(\mathbf{r}_S, \mathbf{r}'_S)$ relates the amplitude of the large and small components of $\psi(\mathbf{r})$ on S and we can use this property to write the current in terms of only the large component of the wavefunction on S,

$$J_S = c^2\hbar\int_S \mathrm{d}^2 r_S \int_S \mathrm{d}^2 r'_S \psi_l^\dagger(\mathbf{r}_S)\Sigma(\mathbf{r}_S, \mathbf{r}'_S)\psi_l(\mathbf{r}'_S). \tag{5.75}$$

We have introduced the 2×2 surface quantity

$$\Sigma(\mathbf{r}_S, \mathbf{r}'_S) = \mathrm{i}\sigma_S\left[\Gamma(\mathbf{r}_S, \mathbf{r}'_S) - \Gamma^\dagger(\mathbf{r}'_S, \mathbf{r}_S)\right]\sigma_{S'} \tag{5.76}$$

where $\sigma_S = \boldsymbol{\sigma}\cdot\mathbf{r}_S$ and $\sigma_{S'} = \boldsymbol{\sigma}\cdot\mathbf{r}'_S$. In the non-relativistic case, Σ becomes the imaginary part of the embedding potential [46]. This is due to the reciprocity of the embedding potential: $\left[G_0^{-1}(\mathbf{r}_S, \mathbf{r}'_S)\right] = \left[G_0^{-1}(\mathbf{r}'_S, \mathbf{r}_S)\right]^*$, but this is not generally true for the relativistic embedding potential because we evaluate the Green's function at energy $W + \mathrm{i}\eta$, which ensures the correct outgoing boundary condition.

The fact that we can write the surface current in terms of the embedding potential is not surprising since we know that Γ describes the band structure of the leads which

contains all the information about the Bloch states incident on S. We must however be careful regarding the direction of the current J_S because of the outgoing boundary condition implicit in the embedding potential. This means that, for the geometry in figure 5-14, J_S is actually the current flowing *out* of the device region Ω *into* the leads.

5.5.3 Channel Functions

We have already seen that the electron states in the semi-infinite leads connected to our device are either extended Bloch states or evanescent surface or interface states which decay into the lead away from the interface. Only the Bloch states carry current in the leads and these are a natural choice of functions for transmission calculations across interfaces [46], [149], [172]. The disadvantage of describing the conduction channels across S_L and S_R by the current carrying Bloch states is that Bloch states are only orthogonal on a surface in the case of free electron leads [46], [117]. For more realistic systems it would be useful to have a set of functions which are automatically orthogonal on S_L and S_R to allow simplification of numerical calculations or to aid physical insight. These are the channel functions and follow from the description of the current in terms of the embedding potential [117].

We can define the eigenfunctions and eigenvalues of $\Sigma(\mathbf{r}_S, \mathbf{r}'_S)$ through

$$\int_S \mathrm{d}^2 r_S \Sigma(\mathbf{r}_S, \mathbf{r}'_S) U_i(\mathbf{r}'_S) = \lambda_i U_i(\mathbf{r}_S). \tag{5.77}$$

The eigenvalues of Σ are real and the eigenfunctions are orthogonal on S because Σ is clearly Hermitian from its definition (5.5.2). The $U_i(\mathbf{r}_S)$ are 2×1 spinors because of the 2×2 nature of Σ and can be identified with the surface values of the large component of a state ψ_i in the lead. The corresponding small component on S can be determined via the embedding potential, and extended into Ω using the Green's function for region Ω

$$\psi(\mathbf{r}) = \mathrm{i}c\hbar \int_S \mathrm{d}^2 \mathbf{r}_S \cdot G(\mathbf{r}, \mathbf{r}_S) \boldsymbol{\alpha} \psi(\mathbf{r}_S), \tag{5.78}$$

so that the U_i define channel functions in the leads. They can be normalised over S

$$\int_S \mathrm{d}^2 r_S U_i^\dagger(\mathbf{r}_S) U_j(\mathbf{r}_S) = \delta_{ij}. \tag{5.79}$$

Multiplying (5.77) from the left by $U_j^\dagger(\mathbf{r}_S)$, integrating over S, using (5.79) and comparing with (5.75) gives the current across S carried by the i^{th} channel function as

$$J_S = c^2 \hbar \lambda_i. \tag{5.80}$$

States with $\lambda_i = 0$ are therefore *closed* and carry no current into the lead, and

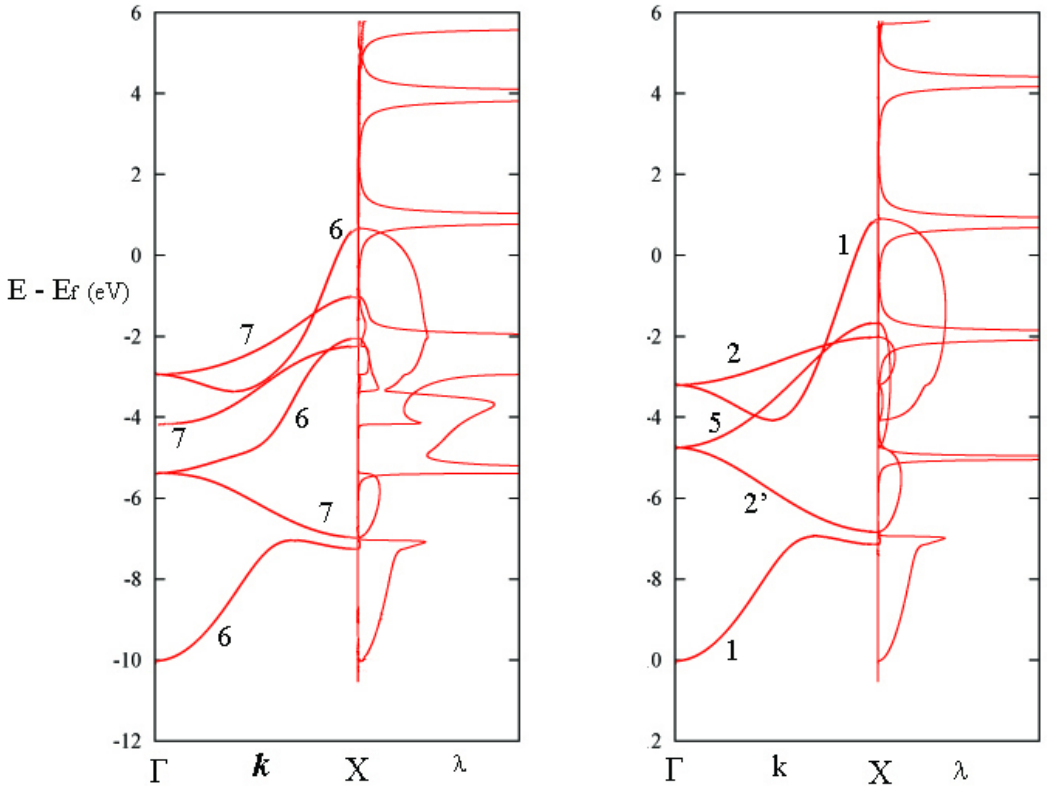

Figure 5-15: Real bands and eigenvalues, λ_i, of Σ for gold along the (001) direction. The left panel shows the fully relativistic bands and eigenvalues, and the right panel those from a scalar relativistic calculation. The bands are labelled by their Δ symmetry.

correspond to evanescent states. Current carrying, *open*, channels have $\lambda_i > 0$ because of the outgoing boundary condition. These open channels are not, however, the same as the Bloch states in the lead, but are related to them by a unitary transformation [117]. This means that the i^{th} channel function and its associated Bloch state carry the same amount of current, or flux, across S. Since the Bloch states can be normalised so that they carry unit flux across S, [173], [174], the open channel functions, with $\lambda_i > 0$, can be chosen so that they also carry unit flux across S:

$$\int_S \mathrm{d}^2 r_S U_i^\dagger(\mathbf{r}_S) U_j(\mathbf{r}_S) = \frac{1}{c^2 \hbar \lambda_i} \delta_{ij}, \quad i \text{ open} \tag{5.81}$$

The real bands and the eigenvalues, λ_i of Σ for gold along the Γ-X (001) direction from a fully relativistic calculation (left) and scalar relativistic calculation (right) are shown in figure 5-15. The scalar relativistic channel function eigenvalues have previously been discussed in ref. [117] and we present the relativistic channel functions for comparison. The calculations were performed with identical parameters to the complex band calculations in section 5.3.3. From figure 5-15 see that each band has an associated eigenvalue, forming a 'bubble' which extends between

the band maximum and minimum. For example, in the non-relativistic case, a large bubble extends from bottom of the topmost Δ_1 band at -4 eV, to its maximum where it crosses the X-point at 0.7 eV. There are also smaller bubbles connecting the band maxima and minima to the points where they cross the Γ-point, such as the one extending from the minimum of the topmost Δ_1 band at -4 eV to where it crosses Γ at -3.2 eV. These bubbles are eigenvalues corresponding to open channel functions associated with current carrying Bloch states. However, the existence of multiple eigenvalues associated with a single band shows that there is not necessarily a one-to-one relationship between Bloch states and channel functions with a given symmetry. These features are also seen in the relativistic channel functions, although the more complicated band structure means they are not perhaps so obvious.

As well as the bubbles corresponding to open channel functions there are delta functions in the spectrum. In order to resolve these a small imaginary part of 4×10^{-6} eV has been added to the energy. These occur when there is a pole in $\Sigma(W)$ at a particular energy, W, and therefore a pole in the embedding potential, $\Gamma(W)$. This means that $\Gamma^{-1}(W)$ has zero eigenvalue and since the embedding potential relates large and small components

$$\psi_s(\mathbf{r}_S) = \mathrm{i}c\hbar \int_S \mathrm{d}^2\mathbf{r}_S \cdot \Gamma(\mathbf{r}_S, \mathbf{r}'_S; W)\boldsymbol{\sigma}\psi_l(\mathbf{r}'_S) \tag{5.82}$$

then the vanishing eigenvalue corresponds to a channel function with a node on S at energy, W. As the imaginary part of the energy tends to zero, the Lorentzian peaks in figure 5-15 become true delta-functions and these zero amplitude wavefunctions carry no current through S.

We will now use the channel functions to derive a relativistic version of the formula, (5.72), for the conductance through a nanodevice.

5.5.4 Conductance

The Landauer approach to ballistic transport is formulated in terms of transmission probabilities for conduction channels across the device region Ω. An initial, incoming state, ϕ, in the left lead L, is transmitted from S_L to S_R across Ω, into an outgoing state χ in the right lead R. The semi-infinite leads can be replaced on S_L and S_R by appropriate embedding potentials and we need only solve the problem explicitly in Ω. We will now show how these transmission probabilities, and hence the conductance, can be determined using channel functions.

The embedding potentials for the leads satisfy the outgoing boundary condition, meaning that they describe states which travel out of Ω and into the leads. The incoming state in the left lead, ϕ, should therefore be thought of as a time reversed outgoing state

$$\phi = \hat{\tau}\psi \tag{5.83}$$

where ψ is a state travelling out of Ω into L and $\hat{\tau}$ is the relativistic time reversal operator introduced in section 2.1.6,

$$\hat{\tau} = \mathrm{i} \begin{pmatrix} \sigma_y & 0 \\ 0 & \sigma_y \end{pmatrix} \tau_0. \tag{5.84}$$

The large component of the incoming state ϕ in the left lead can be extended from S_L through Ω to S_R using the Green's function for Ω, which in terms of the large-large and large-small blocks of the Green's function may be written as

$$\phi_l(\mathbf{r}_R) = -\mathrm{i}c\hbar \int_{S_L} \mathrm{d}^2 r_L \left[G_{ll}(\mathbf{r}_R, \mathbf{r}_L)\sigma_L \phi_s(\mathbf{r}_L) + G_{ls}(\mathbf{r}_R, \mathbf{r}_L)\sigma_L \phi_l(\mathbf{r}_L) \right]. \tag{5.85}$$

The small component of ϕ on S_L is related to the large component via the embedding potential on S_L

$$\phi_s(\mathbf{r}_L) = -\mathrm{i}c\hbar \int_{S_L} \mathrm{d}^2 r'_L \Gamma_L(\mathbf{r}_L, \mathbf{r}'_L)\sigma_L \phi_l(\mathbf{r}'_L) \tag{5.86}$$

and we can also relate the large-large and large-small components of the Green's function in a similar way. To see this we write the Green's function for Ω on the surface S_L in terms of wavefunctions, $\Psi(\mathbf{r})$, valid in Ω

$$G(\mathbf{r}_L, \mathbf{r}'_L) = \sum_{ij} g_{ij} \Psi_i(\mathbf{r}_L) \Psi_j^\dagger(\mathbf{r}'_L) \tag{5.87}$$

then

$$G_{ls}(\mathbf{r}_L, \mathbf{r}'_L) = \sum_{ij} g_{ij} \Psi_{li}(\mathbf{r}_L) \Psi_{sj}^\dagger(\mathbf{r}'_L). \tag{5.88}$$

We can write the small component in terms of the large component using the embedding potential on S_L to give

$$G_{ls}(\mathbf{r}_L, \mathbf{r}'_L) = -\mathrm{i}c\hbar \int_{S_L} \mathrm{d}^2 r''_L G_{ll}(\mathbf{r}_L, \mathbf{r}''_L)\sigma''_L \Gamma_L^\dagger(\mathbf{r}''_L, \mathbf{r}'_L). \tag{5.89}$$

The surface value of the large component of the incoming state in L transferred to S_R is therefore

$$\phi_l(\mathbf{r}_R) = \mathrm{i}c^2\hbar^2 \int_{S_L} \mathrm{d}^2 r_L \int_{S_L} \mathrm{d}^2 r'_L \, G_{ll}(\mathbf{r}_R, \mathbf{r}_L)\Sigma_L(\mathbf{r}_L, \mathbf{r}'_L)\phi_l(\mathbf{r}'_L). \tag{5.90}$$

The incoming state transferred to S_R should match smoothly on to the outgoing states in R, $\chi(\mathbf{r}_R)$ so that $\phi_l(\mathbf{r}_R) = \chi_l(\mathbf{r}_R)$. In deriving (5.90) we considered the incoming state ϕ and its associated embedding potential Γ_L, defined in terms of Bloch and evanescent states propagating and decaying into Ω from out of L. However, embedding potentials appropriate to the embedded region Ω must satisfy the

outgoing boundary condition and be constructed from evanescent and Bloch states propagating and decaying away from S_L into L. We therefore require the time reversed embedding potential and wavefunction in (5.90) to satisfy outgoing boundary conditions for Ω:

$$\chi_l(\mathbf{r}_R) = \mathrm{i}c^2\hbar^2 \int_{S_L} \mathrm{d}^2 r_L \int_{S_L} \mathrm{d}^2 r'_L \, G_{ll}(\mathbf{r}_R, \mathbf{r}_L)\hat{\tau}\Sigma_L(\mathbf{r}_L, \mathbf{r}'_L)\hat{\tau}^\dagger\hat{\tau}\psi_l(\mathbf{r}'_L). \tag{5.91}$$

where $\phi = \hat{\tau}\psi$, and ψ is a wavefunction travelling away from S_L into L.

We are interested in the conductance across Ω and therefore consider ψ_l and χ_l to be the large components of current carrying Bloch states. On the surfaces S_L and S_R these can be expanded in channel functions. For the left lead we consider a single channel i so that

$$\psi(\mathbf{r}_L) = U_{Li}(\mathbf{r}_L). \tag{5.92}$$

which is scattered into the Bloch state in the right lead,

$$\chi_{l,i}(\mathbf{r}_L) = \sum_j t_{ij} U_{Rj}(\mathbf{r}_R) \tag{5.93}$$

where the sum is over open channels, $\lambda_i > 0$ and t_{ij} is a transmission coefficient for scattering from channel i in the left lead to channel j in the right lead. As we have seen, the $U_{Ri}(\mathbf{r}_R)$ are eigenfunctions of Σ_R, constructed from embedding potentials on S_R and the $U_{Li}(\mathbf{r}_L)$ are eigenfunctions of Σ_L constructed from embedding potentials on S_L. The large components of the wavefunctions in L and R are therefore related by

$$\sum_j t_{ij} U_{Rj}(\mathbf{r}_R) = \mathrm{i}c^2\hbar^2 \int_{S_L} \mathrm{d}^2 r_L \int_{S_L} \mathrm{d}^2 r'_L \, G_{ll}(\mathbf{r}_R, \mathbf{r}_L)\hat{\tau}\Sigma_L(\mathbf{r}_L, \mathbf{r}'_L)\hat{\tau}^\dagger\hat{\tau}U_{Li}(\mathbf{r}'_L). \tag{5.94}$$

If we normalise the channel functions so that they carry unit current (5.81), Σ_L can be expanded in its eigenfunctions (the channel functions) as

$$\Sigma_L(\mathbf{r}_L, \mathbf{r}'_L) = c^2\hbar \sum_i \lambda_{Li}^2 U_{Li}(\mathbf{r}_L) U_{Li}^\dagger(\mathbf{r}'_L). \tag{5.95}$$

Substituting into (5.94), multiplying from the left by $U_{Rj}^\dagger(\mathbf{r}_R)$ and integrating over $\mathbf{r}_R$ gives, for the transmission coefficient

$$t_{ij} = \mathrm{i}c^4\hbar^2\lambda_{Rj}\lambda_{Li} \int_{S_R} \mathrm{d}^2 r_R \int_{S_L} \mathrm{d}^2 r_L \, U_{Rj}^\dagger(\mathbf{r}_R) G_{ll}(\mathbf{r}_R, \mathbf{r}_L)\hat{\tau}U_{Li}(\mathbf{r}_L). \tag{5.96}$$

The Landauer formula gives the conductance, C, as a sum over all open channels

of the transmission probabilities between channels

$$C = \sum_{\text{open } ij} |t_{ij}|^2 \tag{5.97}$$

We have

$$\begin{aligned}
|t_{ij}|^2 =& (c^4\hbar^2) \int_{S_R} \mathrm{d}^2 r_R \int_{S_R} \mathrm{d}^2 r'_R \int_{S_L} \mathrm{d}^2 r_L \int_{S_L} \mathrm{d}^2 r'_L \\
& \times \lambda^2_{Li} \left[\hat{\tau} U_{Li}(\mathbf{r}_L)\right]^\dagger G^\dagger_{ll}(\mathbf{r}_R, \mathbf{r}_L) \lambda^2_{Rj} U_{Rj}(\mathbf{r}_R) U^\dagger_{Rj}(\mathbf{r}'_R) G_{ll}(\mathbf{r}'_R, \mathbf{r}'_L) \hat{\tau} U_{Li}(\mathbf{r}'_L) \\
=& (c^4\hbar^2) \int_{S_R} \mathrm{d}^2 r_R \int_{S_R} \mathrm{d}^2 r'_R \int_{S_L} \mathrm{d}^2 r_L \int_{S_L} \mathrm{d}^2 r'_L \\
& \times \mathrm{Tr}\, \hat{\tau} \lambda^2_{Li} U_{Li}(\mathbf{r}'_L) U^\dagger_{Li}(\mathbf{r}_L) \hat{\tau}^\dagger G^\dagger_{ll}(\mathbf{r}_R, \mathbf{r}_L)\, \lambda^2_{Rj} U_{Rj}(\mathbf{r}_R) U^\dagger_{Rj}(\mathbf{r}'_R)\, G_{ll}(\mathbf{r}'_R, \mathbf{r}'_L)
\end{aligned} \tag{5.98}$$

Taking the sum over open channels and using the expansion corresponding to (5.95) for the right hand side gives

$$T = (c^2\hbar^2) \int_{S_R} \mathrm{d}^2 r_R \int_{S_R} \mathrm{d}^2 r'_R \int_{S_L} \mathrm{d}^2 r_L \int_{S_L} \mathrm{d}^2 r'_L\, \mathrm{Tr}\, \hat{\tau} \Sigma_L(\mathbf{r}'_L, \mathbf{r}_L) \hat{\tau}^\dagger\, G^\dagger_{ll}(\mathbf{r}_R, \mathbf{r}_L)\, \Sigma_R(\mathbf{r}_R, \mathbf{r}'_R)\, G_{ll}(\mathbf{r}'_R, \mathbf{r}'_L). \tag{5.99}$$

In the spectral representation, the relativistic Green's function at complex energy W can be written, in terms of eigenfunctions, Ψ_i and eigenvalues W_i of the Dirac equation, as

$$G(\mathbf{r}, \mathbf{r}'; W) = \sum_i \frac{\Psi_i(\mathbf{r}) \Psi^\times(\mathbf{r}')}{W - W_i} \tag{5.100}$$

where

$$\Psi^\times_i(\mathbf{r}) = \left[\hat{\tau} \Psi_i(\mathbf{r})\right]^\dagger. \tag{5.101}$$

It follows from this and the properties of $\hat{\tau}$ that the reciprocity relation for the Green's function is

$$G(\mathbf{r}', \mathbf{r}; W) = \left[\hat{\tau} G(\mathbf{r}, \mathbf{r}'; W) \hat{\tau}^\dagger\right]^\dagger \tag{5.102}$$

and so

$$G^\dagger_{ll}(\mathbf{r}_R, \mathbf{r}_L) = \hat{\tau} G_{ll}(\mathbf{r}_L, \mathbf{r}_R) \hat{\tau}^\dagger. \tag{5.103}$$

The final, relativistic expression for the ballistic conductance across Ω is therefore

$$T = (c^2\hbar^2)^2 \mathrm{Tr}\, \hat{\tau}\, \Sigma_L\, G_{LR}\, \hat{\tau}^\dagger\, \Sigma_R\, G_{LR} \tag{5.104}$$

where we have adopted a simplified notation for clarity, such that $G_{LR} = G_{ll}(\mathbf{r}_L, \mathbf{r}_R)$

etc. and the integrals over S_L and S_R are implied. This expression is independent of the explicit representation of the open, current carrying states in the leads and only requires that each channel carries identical flux, a condition imposed by the normalisation of the channel functions. This constraint is related to the property of the Bloch states in the leads and is an implicit assumption in the Landauer formula.

The non-relativistic expression is

$$T = 4\,\mathrm{Tr}\,\mathrm{Im}\Sigma_L G_{LR}\mathrm{Im}\Sigma_R G^*_{RL}. \tag{5.105}$$

and the first thing to mention when comparing this with the relativistic expression is the appearance of the time reversal operator in the relativistic case. If we look at (5.104) and interpret the Green's function as a propagator, we see that, starting from the left hand side, G_{RL} transfers the large component of the state in the lead L to R across Ω and Σ_R ensures that it has the appropriate small component on S_R. G_{LR} transfers the large component of a state in R through Ω to L and Σ_L enforces the correct small component on S_L. This second part is in the opposite direction to the flow of current, hence the time-reversal operators. We can do the same with the non-relativistic expression, noting that in the case of spin-degenerate non-relativistic systems, the time-reversal operator is simply complex conjugation. We notice, however that the time-reversed Green's function describes propagation from L to R in the direction of the current. However, the conductance T is real and we can complex conjugate (5.105) to give an expression which is directly comparable with the relativistic one. The explicit appearance of the time-reversal operator in the relativistic expression provides an additional level of insight.

In terms of non-relativistic limits, we have already seen that $G_{ll}(\mathbf{r}, \mathbf{r}')$ reduces to the non-relativistic Green's function as $c \to \infty$. The factor of $(c^2\hbar^2)^2$ comes from the fact that the relativistic embedding potential is related to its non-relativistic counterpart by (see section 4.3.1)

$$c^2\hbar^2\sigma_S\Gamma(\mathbf{r}_S, \mathbf{r}'_S)\sigma'_S \to G_0^{-1}(\mathbf{r}_S, \mathbf{r}'_S) \qquad \text{as} \quad c \to \infty. \tag{5.106}$$

For a spin-degenerate system, the embedding potential Γ will be diagonal in spin,

$$\Gamma(\mathbf{r}_S, \mathbf{r}'_S) = \Gamma(\mathbf{r}_S, \mathbf{r}'_S)\mathbb{I}_2 \tag{5.107}$$

and $\Sigma(\mathbf{r}_S, \mathbf{r}'_S)$ reduces to $-2\mathrm{Im}\Gamma(\mathbf{r}_S, \mathbf{r}'_S)$, hence the extra factor of 4 in the non-relativistic expression. However, we also pick up a factor of 2 from the trace in the relativistic case, due to the inclusion of spin, which is not present in the non-relativistic expression.

5.6 Summary

In this chapter we introduced the transfer matrix which relates the boundary values of wavefunctions on the surfaces S and S' on either side of some region Ω. This transfer matrix can be constructed from the Green's function for region Ω with vanishing small components on the surfaces bounding Ω. This Green's function, and hence the transfer matrix, can itself be found from an embedding calculation in Ω with embedding potentials set to zero on the S and S'. The eigenvectors of the transfer matrix for a representative layer of a bulk crystal have been shown to be the boundary values of Bloch and evanescent states on S and S' and the eigenvalues give the complex band structure. The eigenvectors of the transfer matrix provide a means of constructing an embedding potential for a semi-infinite bulk crystal by taking a repeat unit of atoms and introducing additional buffer volumes which transfer the boundary conditions to planar surfaces. These permit simple expansion of the embedding potential and avoid the need to include complicated, curvy embedding surfaces.

The accurate representation of the complex band structure provided by the embedding potential is crucial for transport calculations, as it describes the available conduction channels. The embedding potential allows the definition of a set of functions which, in contrast to the bulk Bloch states, are orthogonal over the surface. These channel functions are ideally suited for investigating transport through thin films or nanodevices connected to semi-infinite leads. We have used the channel functions to derive the relativistic version of a well known formula for the ballistic conductance through such a system, in the Landauer approximation.

Surfaces break the translational symmetry of the bulk crystal and lead to electron states which are localised near the surface. These surface states have wavefunctions which decay away from the surface. The complex band structure, which is 'built in' to the embedding potential calculated from the transfer matrix, contains all the information about these evanescent states as well as the bulk Bloch states, and therefore presents an ideal way of representing the bulk substrate in a surface calculation. In the next chapter we describe how fully relativistic embedded surface calculations can be performed using the full potential linearised augmented plane wave basis.

Chapter 6

EMBEDDED SURFACE CALCULATIONS

In this chapter we present the extension of the ideas developed so far to relativistic embedded surface calculations. We show how these are implemented within the relativistic LAPW framework and describe a demonstration calculation of the spin-orbit splitting of the Au(111) L-gap surface state.

The ideal bulk crystal is a regular arrangement of atoms that extends to infinity in all directions and this regular lattice and infinite translational symmetry means that the potential in which the electrons move is periodic. As a result, the idea of the unit cell, which can be repeated to reproduce the lattice, allows periodic boundary conditions to be imposed and gives rise to Bloch states and the band structure. The band structure enables many physical properties of the bulk to be calculated, for example, most fundamentally, whether the material is a conductor, insulator or semiconductor.

Of course, real crystals cannot extend infinitely, and must have surfaces. In section 4.1 we discussed the wide variety of surface problems to which the non-relativistic embedding method has been applied and, in principle, its relativistic extension can also be applied to the same systems. A type of problem where relativistic effects at surfaces are particularly important is the spin-orbit splitting of surface states. As we have seen in previous chapters, the surface breaks the infinite translational symmetry and the periodic potential is modified near the surface. Far from the surface, the potential remains bulk-like and electron states are delocalised Bloch states with real wavevectors, $\mathbf{k}$. In the surface region, electrons can become 'trapped' by the surface barrier and wavefunctions which decay into the crystal can exist, leading to localised surface states.

Surface states have energies which lie in the forbidden 'energy gaps', which exist within the bulk bands due to the periodic nature of the bulk potential. Away from the surface they may be described by Bloch states with complex wavevec-

tors, $k_z = k_z^R + \mathrm{i}k_z^I$, in the direction normal to the surface. In surface electronic structure calculations, surface states are typically classified as Shockley or Tamm states. Shockley states are free electron-like states whose decaying wavefunctions extend relatively far into the crystal [175], [176]. Tamm states, on the other hand, occur in materials where the bulk electrons are tightly bound, and can be thought of as a 'dangling bond' caused by the absence of a neighbouring atom outside the surface [176].

Just as no real bulk crystal is ideal, impurities and defects will certainly be present, no real surface is a simple termination of a perfectly regular crystal; relaxations and reconstructions will occur. In order for the surface arrangement of atoms to be more stable, the interatomic spacing between layers parallel to the surface may change, or relax. In more extreme cases the lattice may completely reorder itself leading to reconstruction. Some reconstructions can be long-range over the surface and require large surface unit cells in calculations [79], [176]. An additional consideration is whether the surface is 'clean'. The surface leaves atoms in the uppermost layer able to bond with atoms in the free space above and a layer of material may form on the surface, the surface is no longer 'clean'. All of these phenomena can of course modify the surface states and are a major consideration in comparing experiment and calculations. It is obviously easier to perform a calculation on a clean, unreconstructed and unrelaxed surface and in some cases it is possible to prepare experimental samples which are very close to this ideal situation [15], [17].

From the point of view of surface embedding calculations, we need to consider a few things in order to know how to divide the problem into bulk and surface regions. In the presence of a relaxation we need to know how far the change in lattice constant penetrates below the surface and in the case of a reconstruction, how large the unit cell needs to be parallel to the surface. If the surface is not clean, we must include a layer of adsorbed atoms as the uppermost surface layer. In the simplest instance, a clean, unreconstructed or relaxed surface, the most important consideration is how many surface layers we must include in order to include the change in the potential due to the surface. In the case of metals, efficient screening confines the modification of the potential to a few layers below the surface.

So far, we have described how the relativistic bulk and vacuum embedding potentials can be calculated in the LAPW basis and in this chapter we will show how this can be used in a surface calculation. Firstly we will describe the procedure for generating the self-consistent potential in the surface region from the embedded Green's function, paying particular attention to the geometry consisting of the surface layers and buffer regions appropriate for transferring the embedding surface. The final Green's function for the full system can then be constructed, enabling calculation of the surface charge density and density of states. Finally we will present an example calculation for the spin-orbit split surface state on clean Au(111) which has been ex-

tensively investigated experimentally by angle resolved photoemission spectroscopy, and theoretically using slab and KKR calculations, and which therefore permits a detailed assessment of the accuracy of the relativistic embedding method and its implementation.

6.1 The Surface Potential

The non-relativistic embedding method has been applied extensively to surface electronic structure calculations [104], [105], [109], [89], [106], [107], [114], [111], [112], [115], [110], [143] and the extension to relativistic problems is, in general terms, straightforward. The procedure for generating the self-consistent surface potential begins with an initial input charge density, $n_0(\mathbf{r})$, typically chosen to consist of overlapping atomic densities, and its associated potential, $V_0(\mathbf{r})$. From this we find the embedded Green's function, $G(\mathbf{r}, \mathbf{r}'; W)$ for the surface region by inverting $[WO - H + \Gamma_L(W) + \Gamma_R(W)]$ where $\Gamma_L(W)$ and $\Gamma_R(W)$ are the bulk and vacuum embedding potentials. This Green's function is then used to create a new charge density, $n'(\mathbf{r})$, given by

$$n'(\mathbf{r}) = -\frac{1}{\pi} \int \mathrm{d}W \, \mathrm{Im} \, \mathrm{Tr} \, G(\mathbf{r}, \mathbf{r}; W) \tag{6.1}$$

where the integral is over occupied states from below the valence band up to the Fermi energy. The contribution from the core states is added to this charge density, which is calculated by explicit solution in the atomic spheres with the atomic potential. The new charge density is then checked against the initial charge density. If the new charge density differs from the initial density then the two are mixed and the process begins again. When the two are equal we have reached self-consistency and the final surface potential is determined from the final charge density. A schematic view of this self-consistency procedure is shown in figure 6-1. In this section we will describe how this self-consistency cycle is implemented in the relativistic LAPW basis introduced in previous chapters.

In the generation of the bulk embedding potential, we included buffer volumes Δ_L and Δ_R in the embedded region in order to simplify the matrix elements and expansion of the embedding potential by transferring the embedding surface to a plane. The bulk embedding potential was transferred from the curvy surface S_L and expanded on the plane P_L by the inclusion of Δ_L and to compensate for this we must include Δ_L in the surface calculation (see figure 6-2). The embedded region, region I, therefore extends from P_L to the plane P_R on which we apply the vacuum embedding potential. As in the bulk calculation, we define the LAPW wavevectors normal to the surface over the interval $[0 : L]$ so that the basis functions do not satisfy any particular boundary conditions on P_L and P_R. When calculating the bulk embedding potential, sufficient k_n were used in order to provide an accurate

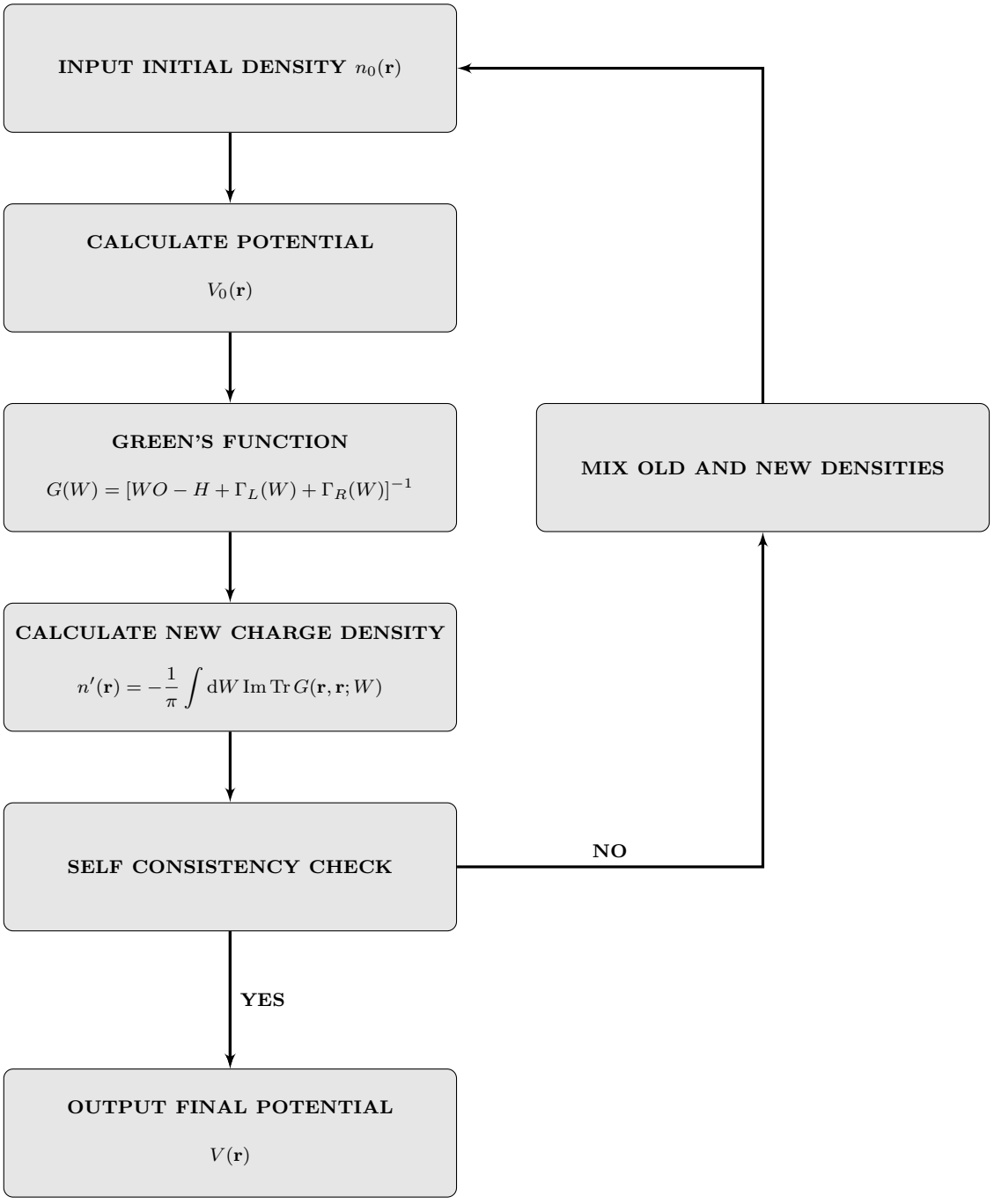

Figure 6-1: Self consistency cycle for generating the surface potential in an embedding calculation.

surface representation of the Green's function used to construct the transfer matrix, corresponding to a cylindrical basis $\{|\mathbf{g}| \leq |\mathbf{G}_{max}|_{||} : k_n = 2\pi n/L, n \leq n_{max}\}$. For the surface calculation, the spherical basis $\{|\mathbf{g}|, k_n \leq |\mathbf{G}_{max}|\}$is usually found to be appropriate.

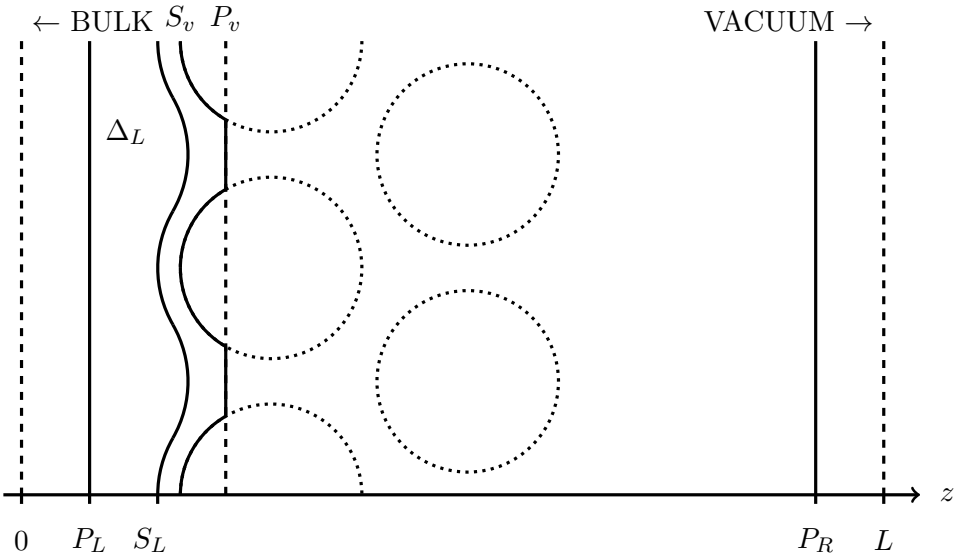

Figure 6-2: Geometry for an embedded surface calculation containing a few surface layers. The bulk embedding potential is transferred from the curvy surface S_L to the plane P_L by including the buffer region Δ_L from the bulk embedding potential calculation. The embedding potential on the vacuum side is expanded on the plane P_R. LAPW basis functions are defined over the interval $[0 : L]$. The plane P_v cuts through the atomic spheres and the 'bumpy' surface S_v follows P_v in the interstitial and the caps of the atomic spheres cut by P_v which project on to the bulk side. These surfaces are used to calculate the Hartree potential.

6.1.1 The Charge Density from the Green's Function

An important part of the surface potential calculation is the generation of the charge density which involves an energy integral of the Green's function $G(\mathbf{r}, \mathbf{r}; W)$. We have frequently mentioned the inclusion of a small imaginary part to the energy $W \to W + \mathrm{i}\eta$ in order to displace the poles of the Green's function from the real energy axis and broaden the delta functions in the density of states at the band energies. this means that the energy integral in the charge density is around a semi-circular contour, C, in the complex plane and we choose a set of complex energies, W_j along this contour. The contour extends from below the bottom of the valence bands up to the Fermi energy, W_f which is defined by the bulk potential calculation and the integral is approximated as

$$\int \mathrm{d}W \,\mathrm{Im}\,\mathrm{Tr}\, G(\mathbf{r}, \mathbf{r}; W) \to \sum_j \mathrm{Im}\,\mathrm{Tr}\, \omega_j G(\mathbf{r}, \mathbf{r}; W_j) \tag{6.2}$$

where the W_j and weights ω_j are generated via a modified Gauss-Chebyshev method [48]. The energies W_j in the Green's function are independent of the pivot energies at which the LAPW basis functions are evaluated, but are identical to the energies of the embedding potentials, $\Gamma(W_j)$ and the same set of energies W_j must be used

in both calculations.

In the LAPW basis, the Green's function itself is calculated by inverting the matrix

$$\left[G(\mathbf{k}_\parallel; W_j)\right]_{\nu\nu'} = \left[\, W_j O(\mathbf{k}_\parallel) - H(\mathbf{k}_\parallel) + \Gamma_L(\mathbf{k}_\parallel; W_j) + \Gamma_R(\mathbf{k}_\parallel; W_j) \,\right]^{-1}_{\nu\nu'} \tag{6.3}$$

where the compound indices denote reciprocal lattice vector, normal wavevector and spin, $\nu = (\mathbf{g}, k_n, m_s)$ for a given parallel wavevector in the 1^{st} two-dimensional Brillouin zone, $\mathbf{k}_\parallel$. To construct the Green's function for the whole embedded region we need to integrate over the Brillouin zone

$$G(\mathbf{r}, \mathbf{r}; W_j) = \int_{BZ} \frac{\mathrm{d}^2\mathbf{k}_\parallel}{(2\pi)^2} G(\mathbf{r}, \mathbf{r}; \mathbf{k}_\parallel, W_j). \tag{6.4}$$

which is again approximated by a summation

$$G(\mathbf{r}, \mathbf{r}; W_j) = \sum_i m_i G(\mathbf{r}, \mathbf{r}; \mathbf{k}_{\parallel_i}, W_j) \tag{6.5}$$

where the m_i are a set of weights (unrelated to those in 6.2). Symmetry is exploited to reduce the number of sampling points required [177]. The Green's function in (6.5) and the matrix elements (6.3) are related by

$$G(\mathbf{r}, \mathbf{r}'; \mathbf{k}_\parallel, W) = \sum_{\nu\nu'} \left[G(\mathbf{k}_{\parallel_i}, W_j)\right]_{\nu\nu'} \psi_\nu(\mathbf{r}; \mathbf{k}_{\parallel_i}) \psi^\dagger_{\nu'}(\mathbf{r}', \mathbf{k}_{\parallel_i}). \tag{6.6}$$

The LAPW basis functions lead to different expansions of the charge density in the interstitial region and atomic spheres. In the interstitial this takes the form

$$n(\mathbf{r}) = \sum_{\mathbf{g},n} n_{\mathbf{g},n} \mathrm{e}^{\mathrm{i}\mathbf{g}\cdot\mathbf{r}_\parallel} \cos(k_n z) \tag{6.7}$$

In the atomic spheres the charge density is expanded as a sum over lattice harmonics

$$n(\mathbf{r}) = \sum_i n_{i\alpha}(|\mathbf{r} - \mathbf{R}_\alpha|) K_{i\alpha}(\mathbf{r} - \mathbf{R}_\alpha). \tag{6.8}$$

To summarise, we first generate and store a bulk embedding potential for an appropriate set of wavevectors and complex energies W_j. These are then used to calculate the Green's function for the embedded region with embedding potentials on the bulk and vacuum sides which are those described in sections 5.4.1 and 5.4.2. For each energy we perform the sum over $\mathbf{k}_{\parallel_i}$ of the trace of the Green's function using (6.5) for the interstitial region and atomic spheres. This is repeated for each energy in the summation approximation to the energy integral around the contour C.

6.1.2 The Potential from the Charge Density

The potential in the surface region is calculated from the charge density in exactly the same way as in non-relativistic embedding calculations. In this section we present the main results from the implementation of Ishida [110].

The embedded surface region has three-dimensional periodicity due to the translational symmetry of the lattice in the direction parallel to the surface and the 'artificial' lattice constant, $2L$, in the z-direction normal to the surface. This means that the exchange correlation potential can be evaluated using the methods employed in standard bulk LAPW calculations [106], [110], [137]. Calculation of the Hartree potential is more subtle because of the buffer region, Δ_L, used to transfer the boundary condition from the 'physical' curvy surface S_L to the unphysical plane P_L on the bulk side.

The charge density in the atomic spheres, consisting of the nuclear charge and the contribution from (6.8), can be replaced by a smooth pseudo charge density, $\tilde{n}(\mathbf{r})$. This pseudocharge density coincides with the 'real' charge density $n(\mathbf{r})$ in the interstitial region and has the same multipole moments as $n(\mathbf{r})$ in the atomic spheres so that it results in the same potential in the interstitial as that generated by the real charge density but allows an expansion in plane waves throughout the embedded region, as in (6.7) [178], [106], [110]. We can now solve the Poisson equation for the potential $\tilde{v}(\mathbf{r})$ over the interval $[P_L : P_R]$

$$\nabla^2 \tilde{v}(\mathbf{r}) = -4\pi \tilde{n}(\mathbf{r}) \tag{6.9}$$

to obtain [110]

$$\tilde{v}_{\mathbf{g}}(z) = \begin{cases} \displaystyle\sum_{n\geq 0} \frac{4\pi \tilde{n}_{\mathbf{g}n}}{k_n^2 + |\mathbf{g}|^2} \cos(k_n z) + V_L \mathrm{e}^{-|\mathbf{g}|(z-P_L)} + V_R \mathrm{e}^{-|\mathbf{g}|(z-P_R)} & \mathbf{g} \neq 0 \\ \displaystyle\sum_{n\geq 1} \frac{4\pi}{k_n^2} \left[\tilde{n}_{\mathbf{g}n} + c_L n_L(z) + c_R n_R(z)\right] \cos(k_n z) & \mathbf{g} = 0 \end{cases}, \tag{6.10}$$

where

$$n_L(z) = \frac{2}{P_L} \sin^2\left(\frac{\pi z}{P_L}\right), \quad n_R(z) = \frac{2}{L - P_R} \sin^2\left[\pi \left(P_R - z\right)\left(L - P_R\right)\right]. \tag{6.11}$$

The coefficients C_L and C_R are chosen so that

$$\left.\frac{\mathrm{d}\tilde{v}_{\mathbf{g}=0}(z)}{\mathrm{d}z}\right|_{P_R} = 0 \tag{6.12}$$

i.e. is flat as it connects with the vacuum potential at P_R. V_R is given by

$$\sum_{n\geq 0} \frac{2\pi\tilde{n}_{\mathbf{g}n}}{k_n^2+|\mathbf{g}|^2}\left[\cos(k_nP_R)-\frac{k_n}{|\mathbf{g}|}\sin(k_nP_R)\right] \tag{6.13}$$

and V_L is matched to the bulk potential on P_L [109], [106].

The potential $\tilde{v}$ matches correctly at the vacuum but not on the bulk side, because we have included the buffer volume Δ_L, and we must add a correction $\phi(\mathbf{r})$ which is a solution of

$$\nabla^2\phi(\mathbf{r})=0 \tag{6.14}$$

so that

$$\tilde{V}_H(S_P)=\phi(S_v)+\tilde{v}(S_v) \tag{6.15}$$

where $\tilde{V}_H(S_v)$ is the Hartree potential corresponding to the pseudo-charge density $\tilde{n}(\mathbf{r})$ in the region between S_v and P_R evaluated on S_v. This Hartree potential, $\tilde{V}_H$ coincides with the real Hartree potential, V_H in the interstitial region. The surface S_v is formed by the plane P_v, which cuts through the atomic spheres, so that in the interstitial S_v is on P_v and when P_v cuts through a sphere, S_v follows the cap of the sphere on the same side of S_v as P_L and is 'bumpy' due to the atomic sphere caps (see figure 6-2).

To deal with the bumpy shape of S_v we first extend the bulk Hartree potential up to S_v through the interstitial region between S_v and P_L and subtract off the 'pseudo' potential $\tilde{v}$,. We then use $V_H(S_v)-\tilde{v}(S_v)$ as the boundary condition for $\phi(S_v)$. For $z\geq P_v$ [110] we expand ϕ as

$$\phi(\mathbf{r})=\sum_{\mathbf{g}}\phi_{\mathbf{g}}\mathrm{e}^{-|\mathbf{g}|(z-P_v)} \tag{6.16}$$

where $\phi_{\mathbf{g}}$ is the $\mathbf{g}$ component of $\phi(\mathbf{r})$ on P_v. If $\mathbf{r}$ is outside the atomic spheres on P_v, i.e. in the interstitial region, then the value of $\phi(P_v)$ is such that it matches to $V_H-\tilde{v}$ on P_v. Inside the spheres the value is unknown and we use an iterative method to find it beginning with some input values of $\phi_{\mathbf{g}}$. In the α^{th} sphere, centred at $\mathbf{R}_\alpha$, we expand $\phi(\mathbf{r})$ in spherical harmonics

$$\phi(\mathbf{r})=\sum_{lm_l}\phi_{lm_l}\left(\frac{|\mathbf{r}-\mathbf{R}_\alpha|}{R_\alpha}\right)^l Y_{lm_l}(\mathbf{r}-\mathbf{R}_\alpha) \tag{6.17}$$

Since we know $\phi(\mathbf{r})$ in the interstitial on the bulk side of S_v we can find the values of $\phi_(\mathbf{r})$ on S_v. To calculate ϕ_{lm_l} we use 6.16) to the right of S_v to get $\phi_{\mathbf{g}}$ and thus determine ϕ_{lm_l}. Using (6.17) we find $\phi(\mathbf{r})$ on the circular part of P_v which intersects the atomic spheres, and we now know $\phi(\mathbf{r})$ on the whole of P_v and have the values of $\phi_{\mathbf{g}}$. We then mix these new $\phi_{\mathbf{g}}$ with the originals and use these as input for a second cycle, continuing until self-consistency. We then have the Hartree potential

$\tilde{V}_H = \tilde{v} + \phi$ corresponding to the pseudo charge density $\tilde{n}(\mathbf{r})$ in the region between S_v and P_R. This coincides with the real Hartree potential, V_H in the interstitial region. The Hartree potential in the atomic spheres can be calculated by integrating the Poisson equation using the atomic potential so that it is continuous at the sphere boundary [137].

The matrix elements of the potential used for the surface calculation are slightly different to those in the generation of the embedding potential. Instead of calculating the matrix elements in the spheres and over the whole of the interstitial region separately and then subtracting off corrections in the atomic spheres, the interstitial potential is tabulated on a real space grid and multiplied by a step function, Θ

$$\Theta = \begin{cases} 1 & \mathbf{r} \quad \text{in the interstitial} \\ \\ 0 & \mathbf{r} \quad \text{in the atomic spheres} \end{cases} . \tag{6.18}$$

The potential can then be expanded in plane waves using a Fast Fourier Transform and the matrix elements are easily evaluated [137].

The self consistency cycle for generating a surface potential can be summarised as follows. An input charge density is chosen, usually that of overlapping atomic charge densities, and from this an initial potential is constructed, which is used to calculate matrix elements of the Hamiltonian, overlap and embedding potentials at a complex energy point W_j and k-point $\mathbf{k}_{\parallel_i}$. These are then used to generate a Green's function and the charge density is calculated by summing over the $\mathbf{k}_{\parallel_i}$ and W_j with the appropriate weights ω_j and w_i. This charge density is then mixed with the input density to generate a new density from which a new potential is calculated over the region between bulk and vacuum embedding planes P_L and P_R. New matrix elements and a new Green's function are calculated, giving a new charge density, and the cycle is repeated until the charge density is self-consistent. A final potential can then be generated to give an output Green's function, used to calculate the surface density of states.

6.2 The Surface Density of States

A key quantity used for investigating surface phenomena is the surface density of states. This is easily determined from the Green's function and is helpful in interpreting, for example, the results of Angle Resolved Photoemission Spectroscopy (ARPES) or scanning tunnelling spectroscopy (STS) experiments. The density of states is obtained from the Green's function for the embedded region as

$$n(W, \mathbf{k}_\parallel) = -\frac{1}{\pi} \int_{\mathbb{V}} \mathrm{d}^3 \, \mathrm{Im} \, \mathrm{Tr} \, G(\mathbf{r}, \mathbf{r}; W, \mathbf{k}_\parallel). \tag{6.19}$$

or more generally can be decomposed into contributions from different regions of space, such as different atoms, or orbital symmetry. In the LAPW basis this becomes, at a single $\mathbf{k}_\parallel$,

$$n(W_j, \mathbf{k}_{\parallel_i}) = -\frac{1}{\pi} \sum_{\nu\nu'} \mathrm{Im} \left[G(\mathbf{k}_{\parallel_i}, W_j) \right]_{\nu\nu'} [O]_{\nu'\nu} \,. \tag{6.20}$$

Depending on the wavevectors or energies at which the density of states is required, it is necessary to pre-calculate an embedding potential at the same values in order to generate the Green's function for the embedded region.

6.2.1 The Spin Polarisation

The spin polarisation can also be obtained from the Green's function in a similar way to the density of states in the LAPW basis:

$$\langle \boldsymbol{\sigma} \rangle = -\frac{1}{\pi} \mathrm{Im} \sum_{\nu\nu'} \left[G(\mathbf{k}_{\parallel_i}, W_j) \right]_{\nu\nu'} [\boldsymbol{\sigma}]_{\nu'\nu} \,. \tag{6.21}$$

where $[\boldsymbol{\sigma}]_{\nu'\nu}$ are matrix elements of the 4×4 spin matrices. This has contributions from the atomic spheres given by

$$[\boldsymbol{\sigma}]_{\nu\nu'} = \frac{\hbar}{2} \sum_{\alpha} \sum_{\Lambda\tau} \sum_{\Lambda'\tau'} (A^{\alpha\tau}_{\Lambda\nu})^* \, A^{\alpha\tau'}_{\Lambda'\nu'} \left[\int_0^{S_\alpha} \mathrm{d}^2\rho \, \rho^2 g_{\kappa\tau}(\rho) g_{\kappa'\tau'}(\rho) \mathbf{I}_{\Lambda\Lambda'} \right.$$

$$\left. + \int_0^{S_\alpha} \mathrm{d}^2\rho \, \rho^2 f_{\kappa\tau}(\rho) f_{\kappa'\tau'}(\rho) \mathbf{I}_{\overline{\Lambda\Lambda'}} \right] \tag{6.22}$$

with $\Lambda = (\kappa, \mu)$, $\tau = 1, 2$,

$$g_{1\kappa}(\rho) = g_\kappa(\rho), \qquad g_{2\kappa}(\rho) = \dot{g}_\kappa(\rho) \quad \text{etc..} \tag{6.23}$$

and

$$\mathbf{I}_{\Lambda\Lambda'} = \int_{4\pi} \mathrm{d}^2\hat{\boldsymbol{\rho}} \, \Omega^\dagger_\Lambda(\boldsymbol{\rho}) \boldsymbol{\sigma} \Omega^\dagger_{\Lambda'}(\boldsymbol{\rho}) \tag{6.24}$$

The spin angular functions can be written in terms of Clebsch-Gordan coefficients, spherical harmonics and two spinors as

$$\Omega_\Lambda(\hat{\boldsymbol{\rho}}) = \sum_{m_s} C_{\Lambda m_s} Y_l^{\mu - m_s}(\hat{\boldsymbol{\rho}}) \phi_{m_s} \tag{6.25}$$

and the required integrals become

$$I^i_{\Lambda\Lambda'} = \delta_{ll'}\Big\{C_{\kappa\mu\frac{1}{2}}C_{\kappa'\mu'\frac{1}{2}}\delta_{\mu\mu'}\sigma^i_{\frac{1}{2}\frac{1}{2}} + C_{\kappa\mu\frac{1}{2}}C_{\kappa'\mu'-\frac{1}{2}}\delta_{\mu\mu'+1}\sigma^i_{\frac{1}{2}-\frac{1}{2}}$$

$$+C_{\kappa\mu-\frac{1}{2}}C_{\kappa'\mu'\frac{1}{2}}\delta_{\mu\mu'-1}\sigma^i_{-\frac{1}{2}\frac{1}{2}} + C_{\kappa\mu-\frac{1}{2}}C_{\kappa'\mu'-\frac{1}{2}}\delta_{\mu\mu'}\sigma^i_{-\frac{1}{2}-\frac{1}{2}}\Big\} \tag{6.26}$$

$$I^i_{\overline{\Lambda}\Lambda'} = \delta_{\overline{l}l'}\Big\{C_{-\kappa\mu\frac{1}{2}}C_{-\kappa'\mu'\frac{1}{2}}\delta_{\mu\mu'}\sigma^i_{\frac{1}{2}\frac{1}{2}} + C_{-\kappa\mu\frac{1}{2}}C_{-\kappa'\mu'-\frac{1}{2}}\delta_{\mu\mu'+1}\sigma^i_{\frac{1}{2}-\frac{1}{2}}$$

$$+C_{-\kappa\mu-\frac{1}{2}}C_{-\kappa'\mu'\frac{1}{2}}\delta_{\mu\mu'-1}\sigma^i_{-\frac{1}{2}\frac{1}{2}} + C_{-\kappa\mu-\frac{1}{2}}C_{-\kappa'\mu'-\frac{1}{2}}\delta_{\mu\mu'}\sigma^i_{-\frac{1}{2}-\frac{1}{2}}\Big\} \tag{6.27}$$

with $i = x, y, z$. The Clebsch-Gordan coefficients are

$$C_{\kappa\mu\frac{1}{2}} = -\mathrm{sgn}(\kappa)\sqrt{\frac{\kappa-\mu+\frac{1}{2}}{2\kappa+1}} \tag{6.28}$$

and

$$C_{\kappa\mu-\frac{1}{2}} = \sqrt{\frac{\kappa+\mu+\frac{1}{2}}{2\kappa+1}}. \tag{6.29}$$

The contribution from the interstitial region is

$$[\boldsymbol{\sigma}]_{\nu\nu'} = \frac{1}{2\mathbb{L}}\delta(\mathbf{g}'-\mathbf{g})$$

$$\Big\{\Big[\phi^\dagger_{m_s}\sigma^i_{m_sm'_s}\phi_{m'_s} + \gamma_k\gamma_{k'}\phi^\dagger_{m_s}\boldsymbol{\sigma}\cdot\mathbf{k}^{(+)}\sigma^i_{m_sm'_s}\boldsymbol{\sigma}\cdot\mathbf{k}'^{(+)}\phi_{m'_s}\Big]\int_{Z_1}^{Z_2}\mathrm{d}z\,\mathrm{e}^{-\mathrm{i}(k_n-k'_n)z}$$

$$-\Big[\phi^\dagger_{m_s}\sigma^i_{m_sm'_s}\phi_{m'_s} + \gamma_k\gamma_{k'}\phi^\dagger_{m_s}\boldsymbol{\sigma}\cdot\mathbf{k}^{(+)}\sigma^i_{m_sm'_s}\boldsymbol{\sigma}\cdot\mathbf{k}'^{(-)}\phi_{m'_s}\Big]\int_{Z_1}^{Z_2}\mathrm{d}z\,\mathrm{e}^{-\mathrm{i}(k_n+k'_n)z}$$

$$-\Big[\phi^\dagger_{m_s}\sigma^i_{m_sm'_s}\phi_{m'_s} + \gamma_k\gamma_{k'}\phi^\dagger_{m_s}\boldsymbol{\sigma}\cdot\mathbf{k}^{(-)}\sigma^i_{m_sm'_s}\boldsymbol{\sigma}\cdot\mathbf{k}'^{(+)}\phi_{m'_s}\Big]\int_{Z_1}^{Z_2}\mathrm{d}z\,\mathrm{e}^{\mathrm{i}(k_n-k'_n)z}$$

$$+\Big[\phi^\dagger_{m_s}\sigma^i_{m_sm'_s}\phi_{m'_s} + \gamma_k\gamma_{k'}\phi^\dagger_{m_s}\boldsymbol{\sigma}\cdot\mathbf{k}^{(-)}\sigma^i_{m_sm'_s}\boldsymbol{\sigma}\cdot\mathbf{k}'^{(-)}\phi_{m'_s}\Big]\int_{Z_1}^{Z_2}\mathrm{d}z\,\mathrm{e}^{\mathrm{i}(k_n-k'_n)z}\Big\}, \tag{6.30}$$

6.3 The Spin Orbit Interaction

A significant motivation for an embedding scheme based on the Dirac equation rather than the scalar relativistic Koelling-Harmon approximation is that it contains the additional relativistic corrections due to the spin orbit interaction. In investigating the consequences of the spin-orbit interaction it is useful to be able to examine the contributions from different angular momentum channels or to switch it off entirely. In the absence of spin orbit effects we would expect the Dirac and Koelling-Harmon approaches to give identical results.

To be able to manipulate (computationally) the spin orbit effects, we need to identify an effective spin orbit term in the Dirac equation. We have already done this when we looked at the non-relativistic limit of the Dirac equation in section 2.1.4. We found that for a spherical potential $V(r)$, retaining terms to order $1/c^2$

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(r) - E\right]\phi(\mathbf{r}) - \frac{\mathbf{p}^4}{8m^3c^2}\phi(\mathbf{r})$$

$$-\frac{\hbar^2}{4m^2c^2}\frac{\mathrm{d}V(r)}{\mathrm{d}r}\frac{\mathrm{d}\phi(\mathbf{r})}{\mathrm{d}r} + \frac{\hbar^2}{4m^2c^2}\frac{1}{r}\frac{\mathrm{d}V(r)}{\mathrm{d}r}\boldsymbol{\sigma}\cdot[\mathbf{r}\times\mathbf{p}]\,\phi(\mathbf{r}) = 0, \tag{6.31}$$

where $\phi(\mathbf{r})$ is the large component of the wavefunction which becomes identical to the non-relativistic wavefunction as $c \to \infty$. If we ignore the momentum correction and the Darwin term we have

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(r) - E\right]\phi(\mathbf{r}) + \frac{\hbar^2}{4m^2c^2}\frac{1}{r}\frac{\mathrm{d}V(r)}{\mathrm{d}r}\boldsymbol{\sigma}\cdot[\mathbf{r}\times\mathbf{p}]\,\phi(\mathbf{r}) \tag{6.32}$$

which is the Schrödinger equation with the spin-orbit term. Writing the large component as

$$\phi(\mathbf{r}) = g_\kappa(r)\Omega_\Lambda(\hat{\mathbf{r}}) \tag{6.33}$$

the spin-orbit term is seen to be

$$\frac{\hbar^2}{4m^2c^2}\frac{1}{r}\frac{\mathrm{d}V(r)}{\mathrm{d}r}\boldsymbol{\sigma}\cdot[\mathbf{r}\times\mathbf{p}]\,g_\kappa(r)\Omega_\Lambda(\hat{\mathbf{r}}). \tag{6.34}$$

where the part in square brackets is $\mathbf{L}$, the angular momentum operator. Recall the K operator encountered in the radial Dirac equation (2.23)

$$K = \boldsymbol{\sigma}\cdot\mathbf{L} + \hbar \quad \Rightarrow \quad \boldsymbol{\sigma}\cdot\mathbf{L} = K - \hbar. \tag{6.35}$$

The eigenfunctions of K are the spin angular functions $\Omega_\Lambda(\hat{\mathbf{r}})$ and

$$K\Omega_\Lambda(\hat{\mathbf{r}}) = -\kappa\Omega_\Lambda(\hat{\mathbf{r}}). \tag{6.36}$$

Therefore, for a spherically symmetric potential $V(r)$, the effective spin orbit term contained in the Dirac equation, to first order in $1/c^2$ is

$$H_{SO} = -\frac{\hbar^2}{4m^2c^2}\frac{1}{r}\frac{\mathrm{d}V(r)}{\mathrm{d}r}\left(\kappa+\hbar\right)g_\kappa(r)\Omega_\Lambda(\hat{\mathbf{r}}). \tag{6.37}$$

The dominant contributions to the spin-orbit interaction originate in the atomic cores, so that we can approximate the total spin orbit effects by summing (6.37) over all atoms in the embedded region. We can also remove the spin orbit interaction from the Dirac equation by subtracting (6.37) and separate out contributions from various angular momentum channels by using the relationship between κ and l,

$$l = \kappa \quad \kappa > 0, \qquad l = -\kappa - 1 \quad \kappa < 0. \tag{6.38}$$

In the LAPW basis, matrix elements of (6.37) are

$$[H_{SO}]_{\nu\nu'} = -\sum_{\Lambda\alpha}\sum_{\tau\tau'}\left(A^{\alpha\tau}_{\Lambda\nu}\right)^* A^{\alpha\tau'}_{\Lambda\nu'}\left(\kappa+\hbar\right)\int_0^{S_\alpha}\mathrm{d}r r^2\, g_{\kappa\tau}(r)g_{\kappa\tau'}(r)\xi(r) \tag{6.39}$$

where

$$\xi(r) = \frac{\hbar^2}{4m^2c^2}\frac{1}{r}\frac{\mathrm{d}V(r)}{\mathrm{d}r} \tag{6.40}$$

6.4 Results: Spin Orbit Splitting of the Au(111) L-Gap Surface State

In order to test the relativistic embedding method, we need to choose a system which demonstrates a clear dependence on relativistic effects. An excellent example is the spin-orbit splitting of the L-gap Shockley surface state on clean, unreconstructed Au(111). We have already mentioned that the embedding method is ideal for the study of surface states because of the treatment of the substrate as a true semi-infinite crystal, allowing an accurate representation of the complex band structure which contains information about the Bloch and evanescent states in the bulk. Furthermore, the embedding method means that surface calculations accurately reproduce surface states without the need for a large number of atomic layers in the surface region.

In section 1.1.2 we saw that the broken inversion symmetry at a surface and spin-orbit coupling can lead to surface states with a Rashba type dispersion

$$E(k) \approx E_0 + \frac{\hbar^2}{2m^*}(k \pm k_0)^2, \quad k_0 = \frac{\hbar^2}{m^*}\alpha_R \tag{6.41}$$

where E_0 is the binding energy, the energy at the Brillouin zone centre, k_0 is related to the Rashba parameter α_R which characterizes the magnitude of the splitting, and m^* is an effective mass. The dispersion of such states therefore consists of two

parabolic bands with a splitting linear in **k**.

These states are of particular interest because of the strong dependence of spin-orbit effects and have applications in spintronics, for example as a means of extracting spin-polarised electrons from non-magnetic materials [179], [180], [181]. A surface state with exactly this characteristic dispersion lies in the energy gap in the bulk band structure at the L-point along Γ-L (in the 111 direction) of Gold. It is a Shockley type surface state arising from mixing (hybridisation) of the s, p and d atomic orbitals at the surface, [179]. Gold is especially interesting because of its widespread use as an electrical contact in nanoscale devices.

6.4.1 Previous Investigations: Experiments and Calculations

The first experimental observation of this spin-orbit induced splitting was by Lashell et. al. who measured the dispersion of the Au(111) L-gap surface state using angle resolved photoemission spectroscopy (ARPES) [11]. They observed a splitting of 110 meV at the Fermi energy, with a splitting in k of $2k_0 = 0.023$ $\mathring{A}^{-1}$ and Fermi level crossings at $k_f = \pm 0.153, \pm 0.176$ $\mathring{A}^{-1}$, and a dispersion characterised by an effective mass of $m^* = 0.25$, and a binding energy of -0.417 meV. The splitting was correctly identified as a result of spin-orbit coupling since a nearly free electron calculation predicts splittings of the order of 10^{-6} eV, yet this ignores the effect of the spin-orbit interaction in the ion cores, where it is strongest. In addition, they considered that the Au(111) surface undergoes a significant reconstruction, but the splitting in k is significantly smaller than the dominant reconstruction reciprocal lattice vector at 0.11 $\mathring{A}^{-1}$.

Subsequent measurements have confirmed these findings and investigations of the spin structure of the surface state have found that the spin-polarisation vector lies in the surface plane and each of the split bands is polarised so that the spins rotate in opposite directions [16], [17], [182], [15], [183]. This is in direct agreement with the spin-polarisation expected for a spin-orbit split surface state discussed in section 1.1.2,

$$P^x_{\pm}(\mathbf{k}_{\parallel}) = \mp\frac{k_y}{|\mathbf{k}_{\parallel}|}, \qquad P^y_{\pm}(\mathbf{k}_{\parallel}) = \pm\frac{k_x}{|\mathbf{k}_{\parallel}|}, \qquad P^z_{\pm}(\mathbf{k}_{\parallel}) = 0. \tag{6.42}$$

Results from an ARPES measurement of the surface state are shown in figure 6-3. Results from other calculations are shown in table 6.4.2 for comparison.

Subsequent to the experimental investigation of this surface state, various theoretical calculations have also been performed. Fully relativistic layer KKR simulations in the atomic sphere approximation reveal a Rashba like dispersion with a binding energy of -0.5 eV, a Fermi level splitting of 0.023 $\mathring{A}^{-1}$ and an effective mass of 0.23 [184]. In these calculations the surface barrier was treated in the LDA approximation with an additional image potential contribution added and

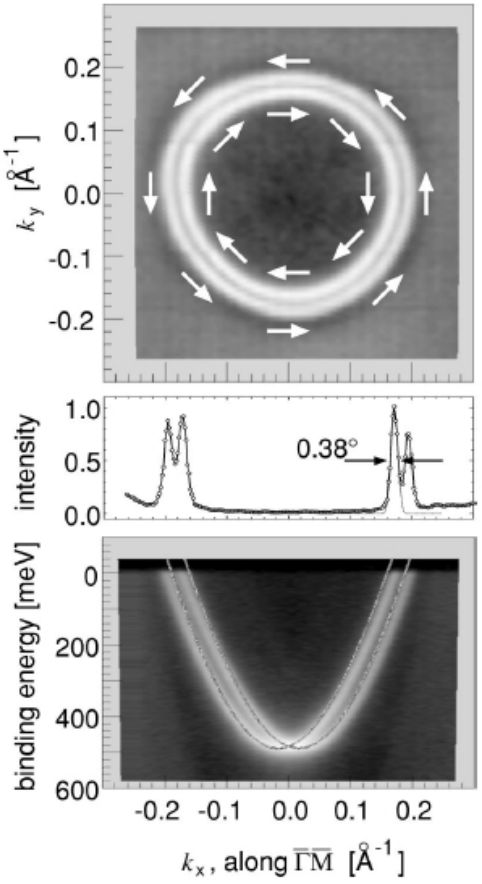

Figure 6-3: Results of an ARPES measurement of the Au(111) surface state (reproduced from [16]. The bottom panel shows the dispersion along k_x in the surface Brillouin zone. The top panel shows the symmetry of the state in the surface Brillouin zone and the arrows show the spin polarisation. The central panel is the momentum distribution of the photo-emitted electrons at $k_y - 0$. The results clearly show the energy splitting of the state.

the bulk is treated as a true semi-infinite crystal. Other semi-infinite KKR simulations, in the muffin tin approximation, give a Fermi level splitting of 0.023 $\mathring{A}^{-1}$ with $k_f = \pm 0.149, \pm 0.172\ \mathring{A}^{-1}$, a binding energy of -0.51 meV and effective mass 0.20 [185], [15]. These were performed with an exchange correlation potential in the LSDA approximation using the Perdew-Wang parameterization. It was found that the interlayer spacing of the surface region needed to be adjusted in order to compensate for the over-steep surface barrier from the LSDA approximation, and to give an accurate work function for the surface and an accurate binding energy for the surface state.

In addition to the semi-infinite substrate calculations, a number of investigations have been carried out using supercell or slab methods using FLAPW basis sets [16], [90], [103], or pseudopotentials [179], again using the LDA and GGA approximations to the exchange correlation potential. The main results from these calculations are

summarised in table 6.4.2 for comparison. A particular characteristic of the slab calculations is the large number of atomic layers employed, 23 or, in the case of [179], 24. This is due to the fact that surface states formed on opposite sides of the slab can interact with each other and cause an energy splitting, even in the absence of the spin orbit interaction. As a result, in order for the fictitious slab-induced splitting to be significantly smaller in energy than the real spin-orbit splitting a large number of layers need to be included in the cell. Forster et. al. [90] found that the artificial splitting is ≈500 meV for 7 layers, ≈5 times the real spin-orbit splitting of 110 meV, and only reduces to ≈ 10 meV when 23 layers are included.

A particularly interesting investigation of the contribution to the splitting from spin orbit effects in different layers was conducted by Bihlmayer et. al. [18]. By using the scalar relativistic approximation as a base and including the spin-orbit interaction only in small spheres around the atomic cores, where it is most significant, they found that only around 58% of the splitting of the Au(111) surface state is due to the spin-orbit interaction in the surface layer, the remaining contributions of 25%, 11% and 4% coming from successively deeper layers. this means that an accurate treatment of the relativistic effects in subsurface layers are required to correctly reproduce the splitting of the surface state.

6.4.2 Embedded Surface Calculations

In this section we will present the results of fully relativistic FLAPW embedded surface calculations of the dispersion and spin polarisation of the L-gap surface state on clean, unreconstructed Au(111). To do this, a previous DFT, scalar relativistic, FLAPW surface embedding code, due to Ishida [110], has been extended to the full Dirac equation using the developments outlined in the preceding chapters of this thesis.

The Density Functional part of the code has not been altered, so that the bulk potential part of the calculation is still performed scalar-relativistically. This potential is used to construct a fully relativistic embedding potential for the substrate. In the generation of the surface potential, relativistic LAPWs have been used to calculate the matrix elements of the charge density and potential, but for the exchange correlation potential we have used the non-relativistic Perdew-Zunger parameterization [58].

In this section we will work in atomic units such that $\hbar = m = e = 1$ and $c = 137.03599$. In these units, energies are measured in Hartrees, Ha, with $1\ Ha = 27.2116\ eV$ and distances in Bohr radii, a_0, $1\ a_0 = 0.52918\ \mathring{A}$. We use an experimental lattice parameter of $4.08\ \mathring{A}$ throughout the calculations.

The bulk potential, consisting of the spherical and non-spherical parts of the potential within the atomic spheres, and the interstitial plane wave potential, is generated from a standard FLAPW calculation and from this calculation we also

determine the Fermi energy. We used a $10 \times 10 \times 10$ Monkhorst-Pack mesh [186] in the irreducible part of the 3-d Brillouin zone with 770 **k**-points and a maximum $|\mathbf{G}|$ of 18.1 a_0^{-1}. For both the embedding potential and surface part of the calculation we find that an atomic sphere radius of 2.63 a_0 is sufficient.

The parameters for the bulk embedding potential used in generating the surface potential were $L = 7.0\ a_0^{-1}$ for the distance over which the k_n are defined, with embedding planes situated at $P_L = 0.7\ a_0^{-1}$ and $P_R = 6.3\ a_0^{-1}$, with the included atomic sphere centred at $z = 3.5\ a_0^{-1}$. If the embedding planes are too close to 0 or L, then it is possible for the basis set to exhibit insufficient freedom to accurately describe the variationally imposed boundary condition. In addition, the embedding planes must also not cut into the atomic spheres, while increasing the z values of the planes can lead to the need for L to increase, meaning a larger number of plane wave basis functions, which decreases the efficiency of the calculation. Within these limits, results for the calculations are independent of the positioning of P_L and P_R in the embedding potential part of the calculation, and P_L in the surface part. The positioning of P_R in the surface region needs to be chosen to accommodate enough space above the surface so that the vacuum level sufficiently constant at P_R. This was achieved if the vacuum was around 7 or 8 a_0 beyond the outermost surface layer.

Next we calculate the surface charge density and potential. Calculations were performed with 1, 2 or 3 atomic layers in the surface region given that the aim is to assess the ability of the embedding scheme to accurately predict surface properties using a relatively small number of surface layers. For each of these situations, we varied the number of **k** points in the two-dimensional surface Brillouin zone, the maximum **g** vector, the number of z-basis vectors, k_n used in constructing the embedding potential and the number of energy points in the energy integral for the charge density. We also investigated the distance above the surface at which the vacuum embedding potential was applied. The accuracy of the calculations was monitored by inspecting the work function, $\phi_0 = V_0 - E_f$ where V_0 is the vacuum level and the Fermi energy, E_f, is carried over from the bulk potential calculation, and by calculating the dispersion of the surface state. The LDA approximation to the exchange correlation potential in the Perdew-Zunger parameterization was used in order to allow comparison with previous studies which have mostly been performed within LDA. Some previous calculations have used GGA approximations but these seem to agree less well with experiment than LDA (see table 6.4.2 for details). In the surface calculation, L took values 16.5 a_0^{-1} for 1 embedded layer, 21 r_b^{-1} for 2 layers and 25 a_0^{-1} for 3 layers. Embedding planes were located at $P_L = 1.5\ a_0^{-1}$ and $P_R = L - 1.5\ a_0^{-1}$ for each calculation.

We find that the surface potential is well converged with an 8×8 mesh with 51 special points, equivalent to 400 **k** points in the surface Brillouin zone. A set of 37 reciprocal lattice vectors up to a cut-off of 3.992 a_0^{-1} were found to be a

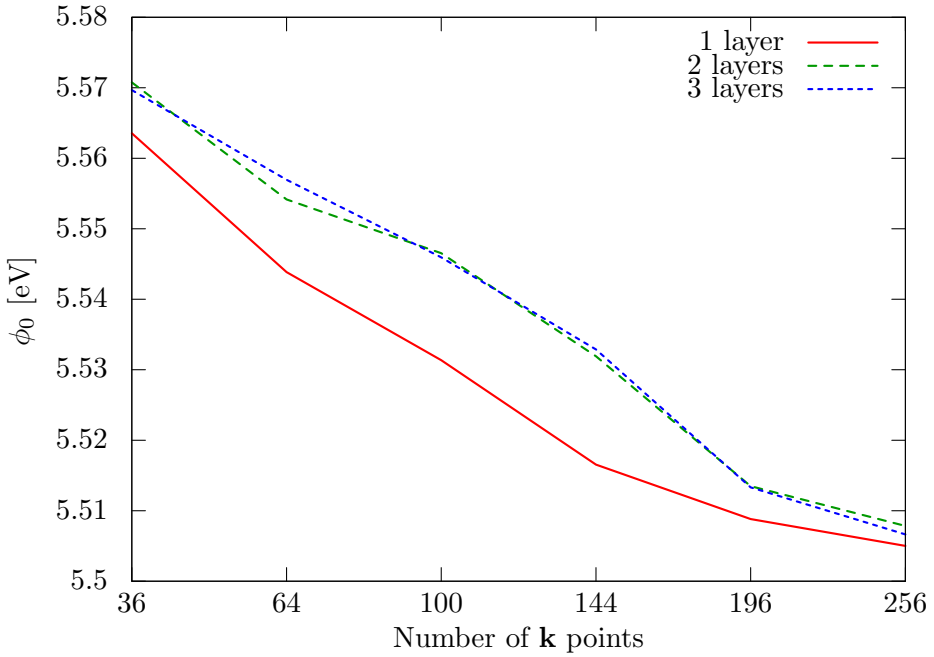

Figure 6-4: The work function of the Au(111) surface calculated with different numbers of **k** points in the surface Brillouin zone. Results of 1, 2 and 3 layer embedded surface calculations are shown.

sufficient basis set in the surface part of the calculation. In the embedding potential calculation, where we use the cylindrical basis, a set of k_n with $n_{max} = 30$ was used. Much larger values of n_{max} can lead to over-completeness unless special steps are taken. For the atomic basis functions, we used a sphere radius of 2.65 a_0^{-1} (≈ 1.4 $\mathring{A}^{-1}$) and an angular momentum cutoff of $l_{max} = 9$. For the energy integral, it was established that no more than 31 Gauss-Chebyshev points are needed. The work functions calculated for different numbers of **k**-points are shown in figure 6-4. The final values, achieved at 256 **k** points are $\phi_0 = 5.505$, 5.507 and 5.508 for 1, 2 and 3 layers respectively, showing excellent consistency and good agreement with other reported values from theoretical studies of the Au(111) surface [179], [185].

Having obtained a self-consistent surface potential and charge density, this was then used to determine the dispersion of the surface state, which lies in the energy gap near the Fermi energy associated with the bulk L-point. A set of calculations over a range of energies and a set of wavevectors along $\overline{\Gamma}$-$\overline{M}$ (k_x) in the surface Brillouin zone with the density of states calculated from the embedded Green's function for the surface region, using an imaginary part of 10^{-4} Hartrees ($\approx 3\times10^{-3}$ eV). Illustrative results are given in figure 6-5, which shows the density of states at $k_x = 0.085$ a_0^{-1}, ≈ 0.16 $\mathring{A}^{-1}$, calculated with 3 atomic layers in the embedded region. The energy splitting of the surface state near -0.1 eV is clear, as is the bulk, and a slight asymmetry in the height of the peaks may be observed. This is due to 'leakage' of the state into the substrate which is energy dependent and greatest when close to the

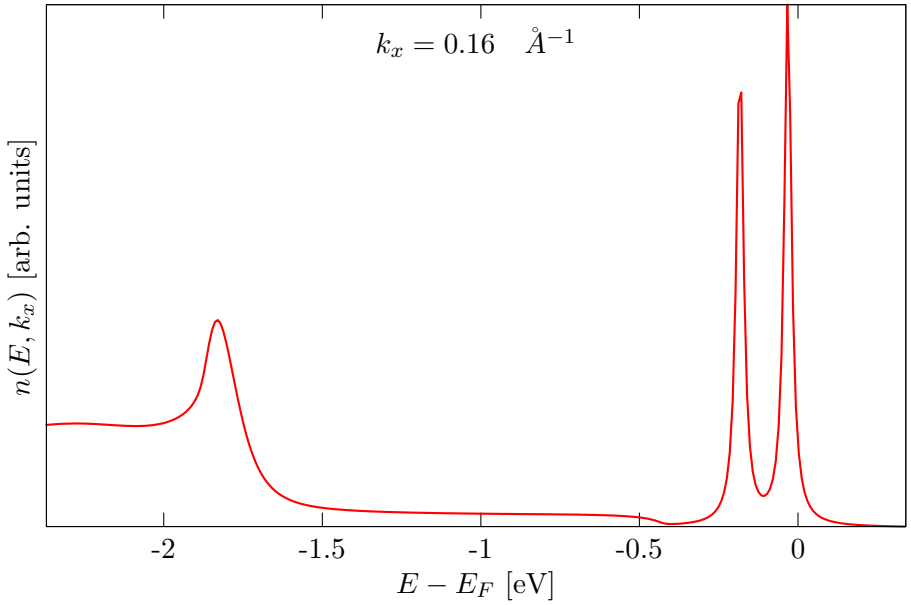

Figure 6-5: Density of states for Au(111) at $k_x = 0.085\ a_0^{-1}$ in the surface Brillouin zone. The L-gap surface state splitting is clearly shown as are the bulk continuum states.

band edge, which is seen at around -0.5 eV and below it is the bulk continuum. This plot is comparable with the energy distribution curve in a photoemission experiment, as shown in the right hand panel of figure 6-7. The fact that the bulk states are represented as a continuum, rather than a discrete spectrum demonstrates that the embedding method treats the substrate as a true semi-infinite crystal.

The full dispersion curve along k_x, for a calculation which contains a single atomic layer, is shown in figure 6-6. Plots such as these are generated from several densities of states, each calculated at a different k_x. The dispersion shows the characteristic split parabola of Rashba-like systems and compares extremely well with results from ARPES measurements (see figures 6-3 and 6-7 and refs [16], [15], [183], [17]). The calculations give binding energies of -0.52 eV for both 1 and 3 surface layers, in good agreement with experiments and previous calculations as may be seen from the compilation of studies summarised in table 6.4.2. It can be seen that theoretical values of the binding energy are typically 20 or 30 meV lower than experimental values, which has previously been attributed to the fact that the LDA approximation does not reproduce the image potential at the surface and leads to a surface barrier which is too steep [185]. However, the one result obtained with the GGA approximation gives a binding energy which is considerably higher, and in much worse agreement with experiment. The effective mass can be extracted from the dispersion curve and we obtain $m^* = 0.25$, a value which agrees with the previous experimental and theoretical results in table 6.4.2. We find that the dispersion curve is not exactly parabolic and the precise value of m^* depends on the range over which

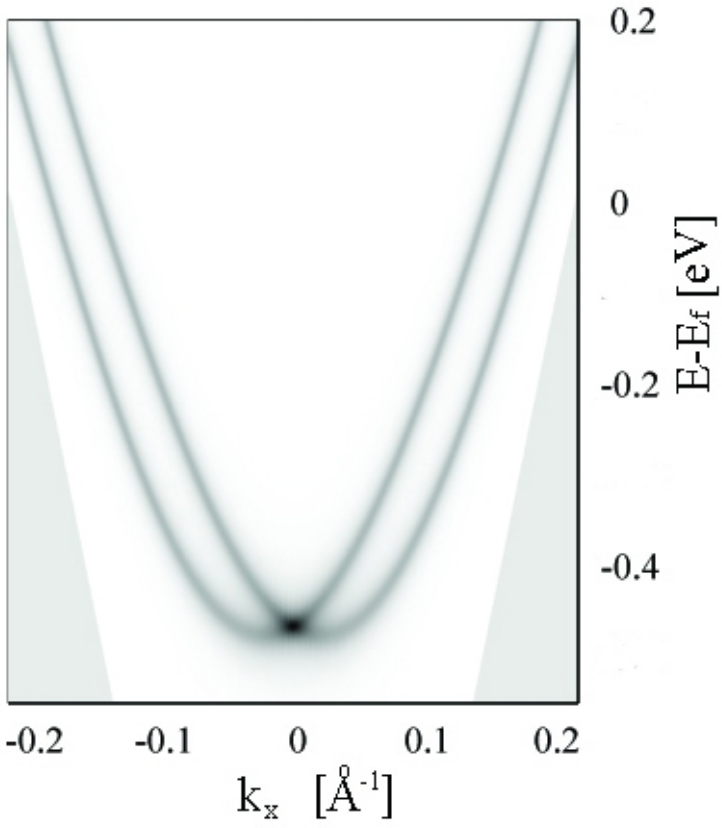

Figure 6-6: Dispersion curve for the Au(111) L-gap surface state from a fully relativistic FLAPW embedded surface calculation. The curve shown is for 1 surface layer and demonstrates the characteristics of a Rashba-split surface state. The shaded areas on either side are the bulk continuum states.

it is calculated.

To demonstrate that the energy splitting of the state is due to the spin orbit interaction, figure 6-8, compares the density of states at k_x=0.095 a_0^{-1} (0.18 $\mathring{A}^{-1}$) from a fully relativistic (bottom) and scalar relativistic (top) calculation. The scalar relativistic approximation contains all relativistic effects except the spin orbit interaction and the absence of any splitting in this calculation clearly demonstrates that the splitting originates from the spin-orbit interaction.

A further means of characterizing the surface state dispersion is by measuring where the surface bands cross the Fermi energy. This is determined by fixing the energy at $E = E_f$ and calculating the density of states for a range of k_x. Figure 6-9 shows results obtained with 1 or 3 atomic layers in the surface region and calculations such as these are directly comparable with momentum distribution curves from photoemission measurements such as that shown in the top panel of figure 6-7. Both 1 and 3 layer calculations give Fermi level crossings at $k_f = \pm 0.155$ and $k_f = \pm 0.184$ $\mathring{A}^{-1}$ and a splitting of $\Delta k_f = 0.029$ $\mathring{A}^{-1}$. These are again in excellent agreement with

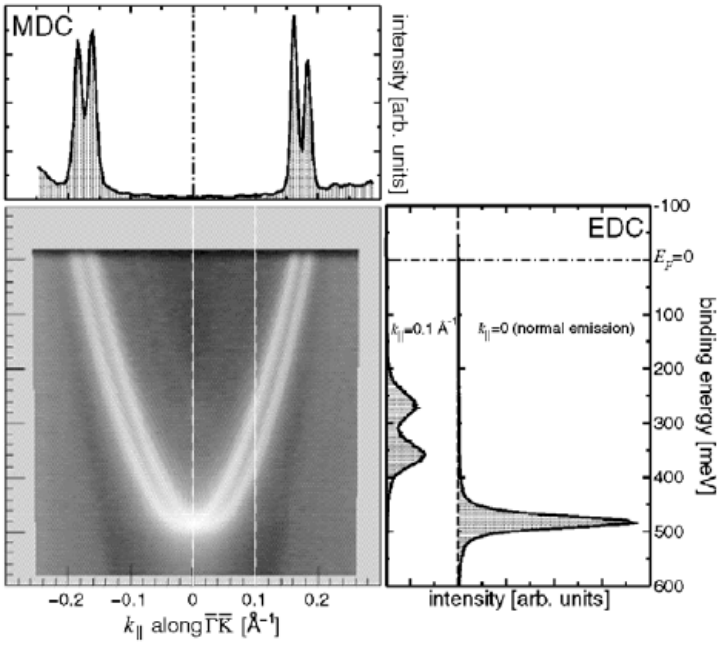

Figure 6-7: Results of a photoemission experiment on the Au(111) surface state. Alongside the dispersion are two different measurements. The top panel shows the momentum distribution of the emitted electrons at a fixed energy, a horizontal slice through the dispersion curve, and the panel on the right hand side shows the energy distribution at a fixed momentum, a vertical slice. Figure reproduced from [17].

previous investigations and follow the trend that theoretical Fermi level splittings are typically $\approx 10\%$ larger than experimental values, which may also be due to the fact that the surface state binding energy is lower because of the steep LDA surface barrier.

An important property of the Rashba-like states at surfaces is the fact that their spin-polarisation lies in the surface plane. This has been observed experimentally [185], [183], [15] and investigated theoretically [185], [184], [15]. We are able to perform a similar study of the spin-polarisation, by calculating the expectation of $\boldsymbol{\sigma}$ in the surface plane as described in section 6.2.1. Calculations show that $\sigma_z = 0$ demonstrating that the spin polarisation does indeed lie in the surface plane. The results of a calculation of $\langle \sigma_y \rangle$ at $E = E_f$ are presented in the topmost panel of figure 6-10 and show that the spin polarisation measured along k_x lies in the k_y direction. The inner state has a spin pointing along $-k_y$ and the outer along $+k_y$ as found previously.

Further insight is gained using the methods outlined in section 6.3, in which the strength of the spin-orbit interaction in different angular momentum channels is varied. There is no contribution of the spin-orbit interaction from the $l = 0$ s-channel, but the p and d channels can be switched on or off separately. To perform such calculations we must calculate an embedding potential which has the relevant

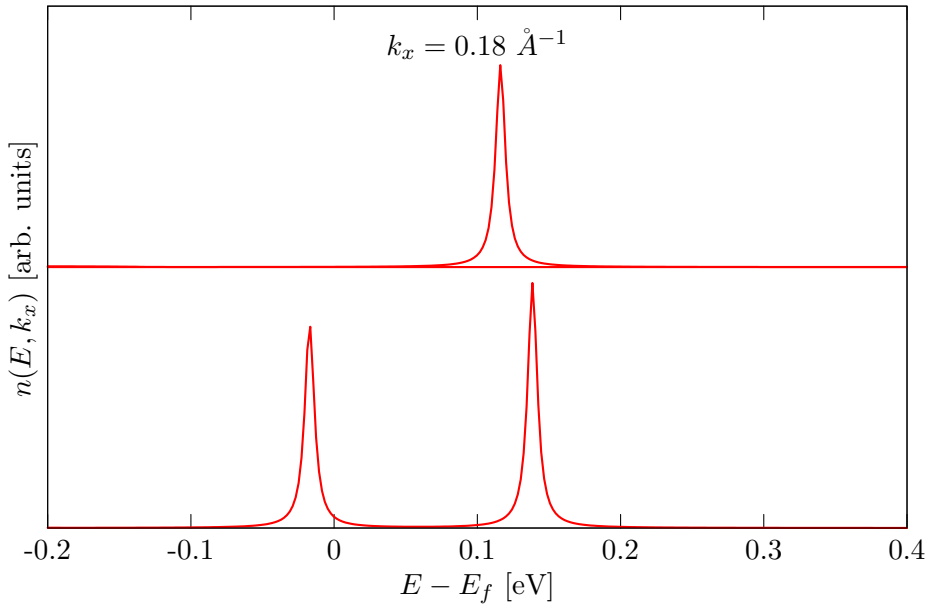

Figure 6-8: Densities of states for Au(111) at $k_x = 0.18$ Å^{-1} from a scalar relativistic (top) and fully relativistic 3 layer embedded calculation, showing the L-gap surface state. The fact that the experimentally observed energy splitting only appears in the fully relativistic calculation demonstrates that it originates from the spin-orbit interaction.

channel switched off in the bulk since a large contribution to the spin splitting of the surface state comes from the subsurface layers. The results of these calculations are also shown in figure 6-10. The upper panel is $\langle \sigma_y \rangle$ with the full spin-orbit interaction, the middle panel has the spin-orbit effects only in the d-channel ($l = 2$) and the lower has it only in the p-channel ($l = 1$). We find essentially no contribution from channels with $l > 2$ (f and above). The larger contribution to the splitting from the d-channel relative to that from the p-channel is in conflict with the results of investigation of the strength of the spin-orbit effect by Nagano et. al [103]. By studying the magnitude of the expectation value of the spin-orbit interaction in the atomic spheres and examining the asymmetry of the surface state wavefunction, they suggested that the contributions from the p-channel was ≈ 8 times larger than that from the d-channel. However, our results suggest that the opposite is true, the p channel contributes less to the overall splitting than the d-channel. From the results shown in figure 6-10, we find that the total wavevector splitting at the Fermi energy is $\Delta k_f = 0.029$ Å^{-1}, from the d-channel it is $\Delta k_f(d) = 0.041$ Å^{-1} and from the p-channel, $\Delta k_f(p) = 0.012$ Å^{-1}. We also find that the polarisation of the spins is reversed in the p-channel and note that $\Delta k_f(d) - \Delta k_f(p) = 0.029$ Å^{-1}.

We can provide a qualitative interpretation of these results by considering the symmetries of the bands forming the top and bottom of the band gap at the L-point in which the surface state exists. The band at the bottom of the gap has p_z character while the band forming the top of the gap is formed by hybridisation of s

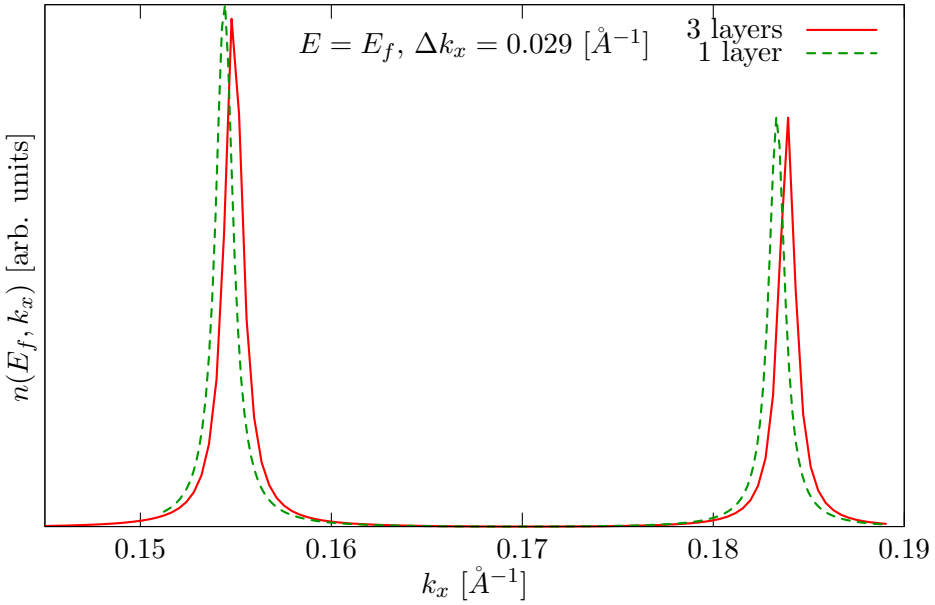

Figure 6-9: Density of states at the Fermi energy showing the wavevector splitting for 1 and 3 embedded surface layers.

and d_{z^2} orbitals and the surface state therefore is *spd*-derived. The reversal in spin polarisation may be explained by considering that the spin orbit interaction involves coupling of the p_z band at the bottom of the gap with p_x and p_y states which are above the band gap and therefore higher in energy. The d_{z^2} band, which forms the bottom of the gap, couples to d-states below the band gap which are lower in energy. From perturbation theory, the term describing the magnitude of the spin-orbit coupling of state $|n\rangle$ to state $|k\rangle$ involves the energy difference between $|n\rangle$ and $|k\rangle$, $E_n - E_k$, hence if $E_n > E_k$, it is positive and for $E_n < E_k$ is negative, providing a possible explanation of the reversal of the spin-polarisation between d and p channels. An investigation of these effects using $k \cdot p$ perturbation theory may provide a better understanding [188].

The fact that we have obtained various results which are in excellent agreement with previous studies demonstrates that the relativistic embedding method is suitable for investigation of the effect of the spin orbit interaction at surfaces. By only including a single layer of surface atoms we can reproduce accurate binding energies and splittings, which shows that the embedding potential faithfully reproduces the properties of the extended bulk substrate. This is particularly true since we know that over 40% of the spin-orbit splitting of the Au(111) surface state comes from layers below the topmost surface layer [18] and the wavefunctions for the surface state extend deep within the substrate, hence the large number of layers required in slab calculations. This fact demonstrates that the embedding potential for the bulk is correctly communicating the spin-orbit effects in the substrate since we achieve accurate results with only 1 layer in the surface region. In addition, the binding

E_0 [eV]	m^*/m_e	k_f[Å^{-1}]	Δk_f[Å^{-1}]	notes
THEORY				
-0.39			0.025	23 layer slab, LDA, FLAPW. [16]
-0.50	0.23		0.023	semi-infinite crystal, LDA+image barrier, atomic sphere [184]
-0.51	0.20	± 0.149, 0.172	0.023	semi-infinite crystal, LDA, muffin tin [185], [15]
-0.484	0.22		0.031	23 layer slab, LDA, FLAPW. [90]
-0.326	0.25		0.031	23 layer slab, GGA, FLAPW. [90]
-0.52	0.24	± 0.159, 0.191	0.032	24 layer slab, LDA, pseudopotential. [179]
			0.028	23 layer slab, FLAPW. [103]
-0.52	0.25	± 0.155, 0.184	0.029	1 embedded layer, LDA, RLAPW.
-0.52	0.25	± 0.155, 0.184	0.029	3 embedded layer, LDA, RLAPW. This work.
EXPERIMENT				
-0.417	0.25	± 0.153, 0.176	0.023	[11]
-0.487	0.255	± 0.172, 0.197	0.025	[16]
-0.487	0.255	± 0.167, 0.192	0.025	[17]
-0.439	0.254	± 0.157, 0.184	0.026	Au(23,23,21). [187]
-0.47	0.25	± 0.160, 0.186	0.026	[15]
-0.46	0.25		0.026	[183]
-0.479	0.26	± 0.172, 0.197	0.024	[90]

Table 6.1: A summary of key parameters from various angle resolved photoemission experimental measurements and theoretical investigations of the Au(111) surface state.

energy of the surface state and surface properties such as the work function demonstrate that the surface potential and charge density can be calculated with sufficient accuracy with the relativistic embedding method.

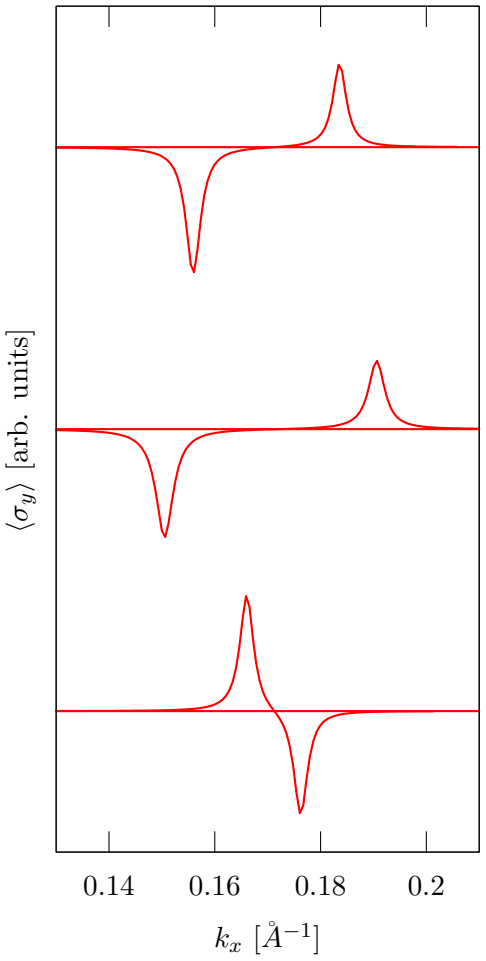

Figure 6-10: Polarisation of the y-component of the spin for the Au(111) surface state at $E = E_f$, calculated from $\langle\sigma_y\rangle$ along k_x with 1 surface layer. The top panel shows the full spin orbit interaction and the spin lies along $\pm k_y$. The middle panel shows the same with the spin-orbit interaction only in the d-channel and the lower only in the p-channel.

Chapter 7

SUMMARY AND OUTLOOK

This thesis has described the extension of the Green's function embedding method of Inglesfield to the Dirac equation, and how this fully relativistic embedding method can be implemented to perform calculations on realistic systems. This has been achieved by adapting an existing full-potential LAPW embedding code in the scalar-relativistic approximation, without spin orbit effects, to relativistic LAPWs based on the full 4-component Dirac equation. By applying this to a well studied problem, we have shown that it produces results which are in excellent agreement with existing experimental and theoretical investigations, along with new insights into underlying mechanisms.

The transfer matrix is commonly used to generate the embedding potential for a semi-infinite substrate in surface calculations and we have derived the relativistic equivalent of this. We have demonstrated that it has similar properties to the non-relativistic transfer matrix and has the correct non-relativistic limit, but also includes contributions which have no non-relativistic analog and therefore are not adequately described by the scalar-relativistic approximation with a first order correction to the spin-orbit interaction. The relativistic transfer matrix is able to provide the complex band structure of the bulk crystal and we have presented results for gold along the (001) and (111) directions alongside scalar relativistic calculations to demonstrate the significance of relativistic corrections due to the spin-orbit interaction in heavy elements.

Since the complex band structure is crucial to understanding electron transport in nanodevices, we have also used our transfer matrix to derive the fully relativistic version of a well known expression for the conductance in these systems. This expression is formally very similar to the non-relativistic version but the fully relativistic treatment provides new insight into its physical interpretation. In investigating transport problems, we have also extended the idea of channel functions to the relativistic case. Non-relativistic channel functions are the eigenfunctions of the imaginary part of the embedding potential and provide an alternative to Bloch states for the study of transport across interfaces. They are better suited to this

because, unlike Bloch states, they are orthogonal over the surface. The relativistic channel functions have the similar properties to their non-relativistic counterparts and we have illustrated this by examining their eigenvalues and comparing with the results of scalar relativistic calculations.

Although we have not described how the relativistic expression for the conductance can be implemented in practice, this has been done for a number of systems in the LAPW basis non-relativistically. Our expression for the conductance, along with the relativistic embedding code, provide an excellent starting point for the fully relativistic investigation of spin transport. In addition, the properties of the channel functions have not been exploited for non-relativistic calculations and since they also arise naturally in the relativistic case this is something which should surely be investigated further.

The strength of non-relativistic embedding is in the calculation of surface electronic structure and we have therefore concentrated on this as the goal of the relativistic extension. We have described how to calculate the potential in the embedded surface region self-consistently using relativistic LAPWs by using scalar-relativistic DFT with the non-relativistic LDA approximation to the exchange correlation potential. An obvious possible development of this is to use fully relativistic density functional methods to obtain this surface potential. Despite the non-relativistic approximation used in generating the surface potential, relativistic calculations of the spin orbit splitting of the dispersion of the Au(111) L-gap surface state yield results which are in excellent agreement with experimental measurements and both non-relativistic and fully relativistic calculations using a variety of methods. We recover comparable binding energies, effective masses, Fermi level crossings, spin polarisations and wavevector splittings. Theoretical values for the binding energy and wavevector splitting at the Fermi energy are typically slightly larger in magnitude due to the LDA approximation which produces a surface barrier which is too steep as it does not correctly reproduce the image potential. A useful extension would be to include this explicitly and to extend the implementation to generalised gradient approximations to the exchange correlation potential.

A further advantage of our implementation for relativistic surface calculations is the ability to calculate the spin-polarisation and decompose this into contributions from the spin-orbit interaction in different angular momentum channels. In terms of the splitting we achieve results which are comparable with previous calculations but we have provided new results for the l-decomposed spin polarisation. These show that the total polarisation consists of contributions from the p and d channels which have opposite ordering of the spin. A provisional argument to explain this reversal of spin polarisation has been presented which presumes that it is due to spin orbit coupling of the components of the surface state in different l channels to bands which are higher or lower in energy relative to the surface state. Quantitative evaluation

of this effect could be achieved by a perturbation theory analysis, giving valuable insight into the mechanisms of spin-orbit splitting of surface states which could be employed in engineering materials for spintronic applications.

Because of the potential applications of Rashba type surface states on non-magnetic metallic surfaces, other systems which exhibit these effects have been sought. Many of these consist of noble metal, simple metal and semi-metal overlayers on metallic or semiconductor substrates in order to enhance the splitting of surface states or manipulate the spin-polarisation [18], [25], [23], [26], [189], [190]. These would be ideal candidates for study with our fully relativistic full-potential LAPW implementation. A further way by which the embedding method is advantageous for systems of this type is that once an embedding potential for a given substrate has been calculated, it can be used again to investigate the effect of adding overlayers on the surface electronic structure. Furthermore, our calculations have shown that only a few layers are needed in the surface region, even for surface states whose wavefunctions extend deep into the crystal, since the embedding potential reproduces the full complex band structure which is crucial for the description of surface states. This is a considerable improvement over slab calculations where the number of layers can be large for such systems and even then they do not faithfully reproduce the bulk continuum states. In addition, the small number of surface layers needed to produce accurate results of the spin-orbit spitting of the surface state shows that the bulk embedding potential correctly reproduces spin-orbit effects in the substrate.

In principle, the relativistic embedding method can be used to study any system for which non-relativistic embedding is suitable. This means that it could be used to calculate Green's functions for investigation of excited states at surfaces, which is particularly useful for their optical response since inclusion of relativistic effects is necessary for a good description of the electron-photon interaction. This idea leads to consideration of an approximation we have used throughout our development of the relativistic embedding method; the absence of a vector potential. With this vector potential, magnetic systems could be studied although it is not immediately obvious how the general embedding method would be changed by this. Clearly, the bulk and surface potentials would need to be calculated in an appropriate DFT framework.

In summary, we have developed a powerful method of determining surface electronic structure which provides a full description of all relativistic effects, including the spin orbit interaction, which is particularly important in a wealth of problems which are currently under intensive study as they are particularly important for the development of a new generation of practical spintronic devices.

Appendix A

MATRIX ELEMENTS

A.1 Basis Functions

The LAPW basis functions in the atomic spheres are

$$\psi_\mu(\mathbf{r}) = \sum_\Lambda \left[A^\alpha_{\Lambda\nu} \begin{pmatrix} g_\kappa(\rho)\Omega_\Lambda(\hat{\boldsymbol{\rho}}) \\ \mathrm{i} f_\kappa(\rho)\Omega_{\overline{\Lambda}}(\hat{\boldsymbol{\rho}}) \end{pmatrix} + B^\alpha_{\Lambda\nu} \begin{pmatrix} \dot{g}_\kappa(\rho)\Omega_\Lambda(\hat{\boldsymbol{\rho}}) \\ \mathrm{i}\dot{f}_\kappa(\rho)\Omega_{\overline{\Lambda}}(\hat{\boldsymbol{\rho}}) \end{pmatrix} \right]. \tag{A.1}$$

where $\boldsymbol{\rho} = \mathbf{r} - \mathbf{R}_\alpha$, R_α is the centre of the α^{th} sphere. The functions $g_\kappa(\rho)$ and $f_\kappa(\rho)$ are solutions of the Dirac equation with the spherically symmetric Coulomb potential at pivot energy w_κ and $\dot{g}_\kappa(\rho)$ and $\dot{f}_\kappa(\rho)$ are their energy derivatives.

In the interstitial region, the basis functions are plane waves

$$\psi_\nu(\mathbf{r}) = \frac{1}{2(\mathrm{i})}\sqrt{\frac{2}{\mathbb{V}}}\mathrm{e}^{\mathrm{i}(\mathbf{k}_\parallel+\mathbf{g})\cdot\mathbf{r}_\parallel} \left[\begin{pmatrix} \phi_{m_s} \\ \gamma_k \boldsymbol{\sigma}\cdot\mathbf{k}^{(+)}\phi_{m_s} \end{pmatrix} \mathrm{e}^{\mathrm{i}k_n z} \pm \begin{pmatrix} \phi_{m_s} \\ \gamma_k \boldsymbol{\sigma}\cdot\mathbf{k}^{(-)}\phi_{m_s} \end{pmatrix} \mathrm{e}^{-\mathrm{i}k_n z} \right]. \tag{A.2}$$

where $\mathbb{V} = \mathbb{A}\mathbb{L}$ is the volume of the unit cell with area $\mathbb{A}$ in the $x-y$ plane and length $\mathbb{L}$ in the z-direction. The composite index $\nu = (\mathbf{k}, m_s)$ and the $\pm$ and the prefactor $1/2(\mathrm{i})$ determine whether the large component is cos-like or sin-like. The ϕ_{m_s} are Pauli 2-spinors,

$$\gamma_k = \frac{c\hbar}{W_k + mc^2}, \qquad W_k = \sqrt{c^2\hbar^2k^2 + m^2c^4}, \tag{A.3}$$

and the wavevector $\mathbf{k} = \mathbf{k}_\parallel + \mathbf{g} \pm k_n$ with kpl a wavevector in the first 2-dimensional Brillouin zone, $\mathbf{g}$ is a 2-dimensional reciprocal lattice vector, The wavevector in the z-direction is

$$k_n = \frac{2\pi n}{L} \tag{A.4}$$

with n even for cos-like and odd for sin-like functions. The length, L, is chosen so that the plane waves do not satisfy particular boundary conditions at the boundaries

of the unit cell; $L > \mathbb{L}$.

A.2 Matching Coefficients

The LAPW coefficients $A_{\Lambda\mu}$ and $B_{\Lambda\mu}$ are determined by matching the basis functions inside the α^{th} atomic sphere at $\mathbf{r} = \mathbf{R}_\alpha$ to the plane waves at the sphere boundary $\rho = S_\alpha$. To do this we must expand the plane waves in spherical waves. The large component can be expanded in relativistic spin angular functions $\Omega_\Lambda(\hat{\mathbf{r}})$ and spherical Bessel functions $j_l(kr)$ by using the identity [10]

$$\phi_{m_s} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} = 4\pi \sum_\Lambda \mathrm{i}^l C_{\Lambda m_s} j_l(kr) \left[Y_l^{\mu - m_s}(\hat{\mathbf{k}}) \right]^* \Omega_\Lambda(\hat{\mathbf{r}}) \tag{A.5}$$

where $C_{\Lambda m_s}$ are Clebsch-Gordan coefficients and $Y_l^{\mu - m_s}$ are spherical harmonics. For the small component we use the relationship between large and small components

$$\psi_s(\mathbf{r}) = \frac{c}{W + mc^2} \boldsymbol{\sigma} \cdot \hat{\mathbf{p}} \psi_l(\mathbf{r}) \tag{A.6}$$

and make use of [10]

$$\boldsymbol{\sigma} \cdot \hat{\mathbf{p}} j_l(kr) \Omega_\Lambda(\hat{\mathbf{r}}) = \mathrm{i}\hbar k \mathrm{sgn}(\kappa) j_{\bar{l}}(kr) \Omega_{\overline{\Lambda}}(\hat{\mathbf{r}}) \tag{A.7}$$

The plane wave expanded in spherical waves is therefore

$$\begin{pmatrix} \phi_{m_s} \\ \gamma_k \boldsymbol{\sigma} \cdot \mathbf{k} \phi_{m_s} \end{pmatrix} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} = 4\pi \sum_\Lambda \mathrm{i}^l C_{\Lambda m_s} \left[Y_l^{\mu - m_s}(\hat{\mathbf{k}}) \right]^* \begin{pmatrix} j_l(kr) \Omega_\Lambda(\hat{\mathbf{r}}) \\ \mathrm{i}\gamma_k k \mathrm{sgn}(\kappa) j_{\bar{l}}(kr) \Omega_{\overline{\Lambda}}(\hat{\mathbf{r}}) \end{pmatrix} \tag{A.8}$$

where

$$\gamma_k = \frac{c\hbar}{W + mc^2}, \qquad W = \sqrt{c^2\hbar^2k^2 + m^2c^4} \tag{A.9}$$

as usual.

The LAPW plane wave basis functions expanded in plane waves are given by

$$\psi(\mathbf{r}) = \sum_\Lambda N_{\nu(\pm)}^{\alpha\Lambda} \begin{pmatrix} j_l(k\rho) \Omega_\Lambda(\hat{\boldsymbol{\rho}}) \\ \mathrm{i}\gamma_k k \mathrm{sgn}(\kappa) j_{\bar{l}}(k\rho) \Omega_{\overline{\Lambda}}(\hat{\boldsymbol{\rho}}) \end{pmatrix} \tag{A.10}$$

with

$$N_{\nu(\pm)}^{\alpha\Lambda} = 2\pi \sqrt{\frac{2}{\mathbb{V}}} \mathrm{i}^{(\pm)} C_{\Lambda m_s} \left\{ \mathrm{e}^{-\mathrm{i}\mathbf{k}^{(+)}\mathbf{R}_\alpha} \left[Y_l^{\mu - m_s}(\hat{\mathbf{k}}^{(+)}) \right]^* \pm \mathrm{e}^{-\mathrm{i}\mathbf{k}^{(-)}\mathbf{R}_\alpha} \left[Y_l^{\mu - m_s}(\hat{\mathbf{k}}^{(-)}) \right]^* \right\} \tag{A.11}$$

where the phase factors come from the radial coordinate being referred to the centre

of the α^{th} atomic sphere at $\mathbf{R}_\alpha$, $\boldsymbol{\rho} = \mathbf{r} - \mathbf{R}_\alpha$. The factor

$$\mathrm{i}^{(\pm)} = \begin{cases} \mathrm{i}^l & \text{for} \quad + \quad \text{(cos-like)} \\ \mathrm{i}^{l-1} & \text{for} \quad - \quad \text{(sin-like)} \end{cases} \tag{A.12}$$

We can now match the atomic solutions to the plane waves at the boundary S_α of the α^{th} atomic sphere. There are two possible matching conditions

A.2.1 Amplitude and Radial Derivative of the Large Component Continuous

This matching condition requires

$$\sum_\Lambda \left[A^\alpha_{\Lambda\nu} g_\kappa(\rho) + B^\alpha_{\Lambda\nu} \dot{g}_\kappa(\rho)\right] \Omega_\Lambda(\hat{\boldsymbol{\rho}}) = \sum_\Lambda N^{\alpha\Lambda}_{\nu(\pm)} j_l(k\rho) \Omega_\Lambda(\hat{\boldsymbol{\rho}})$$

$$\sum_\Lambda \left[A^\alpha_{\Lambda\nu} g'_\kappa(\rho) + B^\alpha_{\Lambda\nu} \dot{g}'_\kappa(\rho)\right] \Omega_\Lambda(\hat{\boldsymbol{\rho}}) = \sum_\Lambda N^{\alpha\Lambda}_{\nu(\pm)} j'_l(k\rho) \Omega_\Lambda(\hat{\boldsymbol{\rho}}), \tag{A.13}$$

where

$$g'_\kappa(\rho) = \frac{\mathrm{d} g_\kappa(\rho)}{\mathrm{d}\rho} \tag{A.14}$$

etc. Multiplying from the left by $\Omega^\dagger_{\Lambda'}(\hat{\boldsymbol{\rho}})$, integrating over $\hat{\boldsymbol{\rho}}$ and using the orthonormality of the spin angular functions

$$\int_{4\pi} \mathrm{d}^2\hat{\boldsymbol{\rho}}\, \Omega^\dagger_{\Lambda'}(\hat{\boldsymbol{\rho}}) \Omega_\Lambda(\hat{\boldsymbol{\rho}}) = \delta_{\Lambda\Lambda'} \tag{A.15}$$

we have

$$A^\alpha_{\Lambda\nu} g_\kappa(S_\alpha) + B^\alpha_{\Lambda\nu} \dot{g}_\kappa(S_\alpha) = N^{\alpha\Lambda}_{\nu(\pm)} j_l(kS_\alpha)$$

$$A^\alpha_{\Lambda\nu} g'_\kappa(S_\alpha) + B^\alpha_{\Lambda\nu} \dot{g}'_\kappa(S_\alpha) = N^{\alpha\Lambda}_{\nu(\pm)} j'_l(kS_\alpha), \tag{A.16}$$

with

$$g'_\kappa(S_\alpha) = \left.\frac{\mathrm{d} g_\kappa(S_\alpha)}{\mathrm{d}\rho}\right|_{S_\alpha}. \tag{A.17}$$

Rearranging gives expression for the LAPW coefficients

$$A^\alpha_{\Lambda\nu} = \frac{N^{\alpha\Lambda}_{\nu(\pm)}}{\Delta^{(1)}_{\kappa\alpha k}} \left[\dot{g}'_\kappa(S_\alpha) j_l(kS_\alpha) - \dot{g}_\kappa(S_\alpha) j'_l(kS_\alpha)\right]$$

$$B^\alpha_{\Lambda\nu} = \frac{N^{\alpha\Lambda}_{\nu(\pm)}}{\Delta^{(1)}_{\kappa\alpha k}} \left[g_\kappa(S_\alpha) j'_l(kS_\alpha) - g'_\kappa(S_\alpha) j_l(kS_\alpha)\right] \tag{A.18}$$

where

$$\Delta^{(1)}_{\kappa\alpha k} = g_\kappa(S_\alpha)\dot{g}'_\kappa(S_\alpha) - \dot{g}_\kappa(S_\alpha)g'_\kappa(\S_\alpha). \tag{A.19}$$

A.2.2 Amplitude of Large and Small Components Continuous

Once again we multiply from the left by $\Omega^\dagger_{\Lambda'}(\hat{\boldsymbol{\rho}})$, integrating over $\hat{\boldsymbol{\rho}}$ and use the orthonormality of the spin angular functions to get

$$A^\alpha_{\Lambda\nu} = \frac{N^{\alpha\Lambda}_{\nu(\pm)}}{\Delta^{(2)}_{\kappa\alpha k}}\left[\dot{f}_\kappa(S_\alpha)j_l(kS_\alpha) - \gamma_k k\mathrm{sgn}(\kappa)\dot{g}_\kappa(S_\alpha)j_{\bar{l}}(kS_\alpha)\right]$$

$$B^\alpha_{\Lambda\nu} = \frac{N^{\alpha\Lambda}_{\nu(\pm)}}{\Delta^{(2)}_{\kappa\alpha k}}\left[\gamma_k k\mathrm{sgn}(\kappa)g_\kappa(S_\alpha)j_{\bar{l}}(kS_\alpha) - f_\kappa(S_\alpha)j_l(kS_\alpha)\right] \tag{A.20}$$

where

$$\Delta^{(2)}_{\kappa\alpha k} = g_\kappa(S_\alpha)\dot{f}_\kappa(S_\alpha) - \dot{g}_\kappa(S_\alpha)f_\kappa(\S_\alpha). \tag{A.21}$$

A.3 The Overlap Matrix

The overlap matrix is

$$[O]_{\nu\nu'} = \int_{\mathbb{V}} \mathrm{d}^3\mathbf{r}\,\psi^\dagger_\nu(\mathbf{r})\psi_{\nu'}(\mathbf{r}) \tag{A.22}$$

where $\mathbb{V}$ is the unit cell volume

$$\mathbb{V} = \mathbb{A}\mathbb{L} \tag{A.23}$$

with $\mathbb{A}$ the area of the 2-dimensional cell and $\mathbb{L}$ the length in the z-direction, from $z = Z_1$ to $z = Z_2$.

Calculation of the overlap matrix is divided into three parts; the overlap over the whole cell, $\mathbb{V}$ using the interstitial basis functions, the overlap in the atomic spheres using the atomic basis functions and finally a correction where we subtract the interstitial overlap from the spheres

$$[O]_{\nu\nu'} = [O_{int}]_{\nu\nu'} + [O_{sph}]_{\nu\nu'} - [O_{corr}]_{\nu\nu'} \tag{A.24}$$

A.3.1 Overlap in the Interstitial Region

In the interstitial we evaluate the overlap integral with the plane wave basis functions.

$$
\begin{aligned}
[O_{int}]_{\nu\nu'} = \frac{1}{2\mathbb{V}} \int_{\mathbb{A}} \mathrm{d}^2\mathbf{r}_{\parallel}\, \mathrm{e}^{\mathrm{i}(\mathbf{g}'-\mathbf{g})\cdot\mathbf{r}_{\parallel}} \\
\times \int_{Z_1}^{Z_2} \mathrm{d}z \left[\begin{pmatrix} \phi_{m_s} \\ \gamma_k \boldsymbol{\sigma}\cdot\mathbf{k}^{(+)}\phi_{m_s} \end{pmatrix} \mathrm{e}^{\mathrm{i}k_n z} \pm \begin{pmatrix} \phi_{m_s} \\ \gamma'_k \boldsymbol{\sigma}\cdot\mathbf{k}'^{(-)}\phi_{m_s} \end{pmatrix} \mathrm{e}^{-\mathrm{i}k'_n z} \right]^{\dagger} \\
\times \left[\begin{pmatrix} \phi_{m_s} \\ \gamma_k \boldsymbol{\sigma}\cdot\mathbf{k}'^{(+)}\phi_{m_s} \end{pmatrix} \mathrm{e}^{\mathrm{i}k'_n z} \pm \begin{pmatrix} \phi_{m_s} \\ \gamma'_k \boldsymbol{\sigma}\cdot\mathbf{k}'^{(-)}\phi_{m_s} \end{pmatrix} \mathrm{e}^{-\mathrm{i}k'_n z} \right],
\end{aligned}
\tag{A.25}
$$

which reduces to

$$
\begin{aligned}
[O_{int}]_{\nu\nu'} = \frac{1}{2\mathbb{L}} \delta(\mathbf{g}'-\mathbf{g}) \\
\left\{ \left[\delta_{m_s m'_s} + \gamma_k \gamma_{k'} \phi^{\dagger}_{m_s} \boldsymbol{\sigma}\cdot\mathbf{k}^{(+)} \boldsymbol{\sigma}\cdot\mathbf{k}'^{(+)} \phi_{m'_s} \right] \int_{Z_1}^{Z_2} \mathrm{d}z\, \mathrm{e}^{-\mathrm{i}(k_n - k'_n)z} \right. \\
- \left[\delta_{m_s m'_s} + \gamma_k \gamma_{k'} \phi^{\dagger}_{m_s} \boldsymbol{\sigma}\cdot\mathbf{k}^{(+)} \boldsymbol{\sigma}\cdot\mathbf{k}'^{(-)} \phi_{m'_s} \right] \int_{Z_1}^{Z_2} \mathrm{d}z\, \mathrm{e}^{-\mathrm{i}(k_n + k'_n)z} \\
- \left[\delta_{m_s m'_s} + \gamma_k \gamma_{k'} \phi^{\dagger}_{m_s} \boldsymbol{\sigma}\cdot\mathbf{k}^{(-)} \boldsymbol{\sigma}\cdot\mathbf{k}'^{(+)} \phi_{m'_s} \right] \int_{Z_1}^{Z_2} \mathrm{d}z\, \mathrm{e}^{\mathrm{i}(k_n - k'_n)z} \\
\left. + \left[\delta_{m_s m'_s} + \gamma_k \gamma_{k'} \phi^{\dagger}_{m_s} \boldsymbol{\sigma}\cdot\mathbf{k}^{(-)} \boldsymbol{\sigma}\cdot\mathbf{k}'^{(-)} \phi_{m'_s} \right] \int_{Z_1}^{Z_2} \mathrm{d}z\, \mathrm{e}^{\mathrm{i}(k_n - k'_n)z} \right\},
\end{aligned}
\tag{A.26}
$$

where

$$
\begin{aligned}
\boldsymbol{\sigma}\cdot\mathbf{k}^{(\pm)} \boldsymbol{\sigma}\cdot\mathbf{k}'^{(\pm)} = \\
\begin{pmatrix} (\pm)k_n(\pm)k'_n + (k_x - \mathrm{i}k_y)(k'_x + \mathrm{i}k'_y) & (\pm)k_n(k'_x - \mathrm{i}k'_y) - (k_x - \mathrm{i}k_y)(\pm)k_n \\ (\pm)k_n(k'_x + \mathrm{i}k'_y) - (k_x + \mathrm{i}k_y)(\pm)k_n & (\pm)k_n(\pm)k'_n + (k_x - \mathrm{i}k_y)(k'_x + \mathrm{i}k'_y) \end{pmatrix}
\end{aligned}
\tag{A.27}
$$

and the z-integrals are

$$\int_{Z_1}^{Z_2} \mathrm{d}z\, \mathrm{e}^{\pm \mathrm{i}(k_n \pm k'_n)z} = \begin{cases} \dfrac{1}{\pm \mathrm{i}(k_n \pm k'_n)} \left[\mathrm{e}^{\pm \mathrm{i}(k_n \pm k'_n)Z_2} - \mathrm{e}^{\pm \mathrm{i}(k_n \pm k'_n)Z_1} \right] & k_n \pm k'_n \neq 0 \\ Z_2 - Z_1 & k_n \pm k'_n = 0 \end{cases} \tag{A.28}$$

A.3.2 Overlap in the Atomic Spheres

In the spheres, the overlap is

$$\begin{aligned} [O_{sph}]_{\nu\nu'} = \sum_{\alpha} \sum_{\Lambda\Lambda'} \int_0^{S_\alpha} \mathrm{d}\rho\, \rho^2 \int_{4\pi} \mathrm{d}^2\hat{\boldsymbol{\rho}} \\ \times \left[A^{\alpha}_{\Lambda\nu} \begin{pmatrix} g_\kappa(\rho)\Omega_\Lambda(\hat{\boldsymbol{\rho}}) \\ \mathrm{i} f_\kappa(\rho)\Omega_{\overline{\Lambda}}(\hat{\boldsymbol{\rho}}) \end{pmatrix} + B^{\alpha}_{\Lambda\nu} \begin{pmatrix} \dot{g}_\kappa(\rho)\Omega_\Lambda(\hat{\boldsymbol{\rho}}) \\ \mathrm{i} \dot{f}_\kappa(\rho)\Omega_{\overline{\Lambda}}(\hat{\boldsymbol{\rho}}) \end{pmatrix} \right]^\dagger \\ \times \left[A^{\alpha}_{\Lambda'\nu'} \begin{pmatrix} g_{\kappa'}(\rho)\Omega_{\Lambda'}(\hat{\boldsymbol{\rho}}) \\ \mathrm{i} f_{\kappa'}(\rho)\Omega_{\overline{\Lambda}'}(\hat{\boldsymbol{\rho}}) \end{pmatrix} + B^{\alpha}_{\Lambda'\nu'} \begin{pmatrix} \dot{g}_{\kappa'}(\rho)\Omega_{\Lambda'}(\hat{\boldsymbol{\rho}}) \\ \mathrm{i} \dot{f}_{\kappa'}(\rho)\Omega_{\overline{\Lambda}'}(\hat{\boldsymbol{\rho}}) \end{pmatrix} \right]. \end{aligned} \tag{A.29}$$

The integral over solid angles makes use of the orthogonality of the spin angular functions

$$\int_{4\pi} \mathrm{d}^2\hat{\boldsymbol{\rho}}\, \Omega^\dagger_\Lambda(\hat{\boldsymbol{\rho}})\Omega_{\Lambda'}(\hat{\boldsymbol{\rho}}) = \delta_{\Lambda\Lambda'} \tag{A.30}$$

and a further simplification is made if the radial functions are normalised so that

$$\int_0^{S_\alpha} \mathrm{d}\rho\, \rho^2 \left[g^2_\kappa(\rho) + f^2_\kappa(\rho) \right] = 1 \tag{A.31}$$

and

$$\int_0^{S_\alpha} \mathrm{d}\rho\, \rho^2 \dot{g}_\kappa(\rho) g_\kappa(\rho) = \int_0^{S_\alpha} \mathrm{d}\rho\, \rho^2 \dot{f}_\kappa(\rho) f_\kappa(\rho) = 0. \tag{A.32}$$

With this normalisation, the overlap in the spheres is

$$[O_{sph}]_{\nu\nu'} = \sum_{\alpha} \left\{ A^{\alpha *}_{\Lambda\nu} A^{\alpha}_{\Lambda\nu'} + F_{\kappa\alpha} B^{\alpha *}_{\Lambda\nu} B^{\alpha}_{\Lambda\nu'} \right\} \tag{A.33}$$

where

$$F_{\kappa\alpha} = \int_0^{S_\alpha} \mathrm{d}\rho\, \rho^2 \left[\dot{g}^2_\kappa(\rho) + \dot{f}^2_\kappa(\rho) \right]. \tag{A.34}$$

A.3.3 Interstitial Correction in the Atomic Spheres

We need to subtract the interstitial overlap from the atomic spheres. We do this by using the spherical wave expansion of the plane wave basis functions. In this case the overlap correction is

$$[O_{corr}]_{\nu\nu'} = \sum_{\alpha} \sum_{\Lambda\Lambda'} \int_0^{S_\alpha} \mathrm{d}\rho\, \rho^2 \int_{4\pi} \mathrm{d}^2\hat{\boldsymbol{\rho}}\, N^{\alpha\Lambda *}_{\nu(\pm)}\, N^{\alpha\Lambda'}_{\nu'(\pm)}$$

$$\times \begin{pmatrix} j_l(k\rho)\Omega_\Lambda(\hat{\boldsymbol{\rho}}) \\ \mathrm{i}\gamma_k k \mathrm{sgn}(\kappa) j_{\bar{l}}(k\rho)\Omega_{\bar{\Lambda}}(\hat{\boldsymbol{\rho}}) \end{pmatrix}^\dagger \begin{pmatrix} j_{l'}(k'\rho)\Omega_{\Lambda'}(\hat{\boldsymbol{\rho}}) \\ \mathrm{i}\gamma_{k'} k' \mathrm{sgn}(\kappa') j_{\bar{l}'}(k'\rho)\Omega_{\bar{\Lambda}'}(\hat{\boldsymbol{\rho}}) \end{pmatrix} \tag{A.35}$$

We can use the orthogonality of the spin angular functions again and the correction reduces to

$$[O_{corr}]_{\nu\nu'} = \sum_{\alpha} N^{\alpha\Lambda *}_{\nu(\pm)}\, N^{\alpha\Lambda}_{\nu'(\pm)} \int_0^{S_\alpha} \mathrm{d}\rho\, \rho^2 \left\{ j_l(k\rho) j_l(k'\rho) + \gamma_k \gamma_{k'} k k' j_{\bar{l}}(k\rho) j_{\bar{l}}(k'\rho) \right\} \tag{A.36}$$

A.4 Hamiltonian Matrix Elements

A.4.1 Kinetic Energy Matrix Elements in the Interstitial Region

The general expression for matrix elements of the embedded Hamiltonian in the interstitial can be separated into contributions from the kinetic energy and potential

$$[H_{int}]_{\nu\nu'} = [T]_{\nu\nu'} + [V]_{\nu\nu'} \tag{A.37}$$

The kinetic energy is

$$[T]_{\nu\nu'} = \int_{\mathbb{V}} \mathrm{d}^3\mathbf{r} \psi^\dagger_\nu(\mathbf{r}) \left\{ -\mathrm{i}c\hbar\boldsymbol{\alpha}\cdot\nabla + \beta mc^2 \right\} \psi_{\nu'}(\mathbf{r}) + \mathrm{i}c\hbar \int_S \mathrm{d}^2\mathbf{r}_S \cdot \psi^\dagger_{l\,\nu}(\mathbf{r}_S)\boldsymbol{\sigma}\psi_{s\,\nu'}(\mathbf{r}_S), \tag{A.38}$$

where the surface integral comes from the discontinuity of the small component of the wavefunction on S.

We can simplify the calculation of these matrix elements as follows. The term in βmc^2 is

$$\int_{\mathbb{V}} \mathrm{d}^3\mathbf{r} \psi^\dagger_\nu(\mathbf{r}) \beta mc^2 \psi_{\nu'}(\mathbf{r}) = mc^2 \left([O]^{ll}_{\nu\nu'} - [O]^{ss}_{\nu\nu'} \right) \tag{A.39}$$

where $[O]^{ll}_{\nu\nu'}$ and $[O]^{ss}_{\nu\nu'}$ are the large-large and small-small components of the overlap matrix. The term in $\boldsymbol{\alpha}\cdot\nabla$ can be expanded in large and small components in a similar

way

$$-\mathrm{i}c\hbar\int_{\mathbb{V}}\mathrm{d}^3\mathbf{r}\psi_\nu^\dagger(\mathbf{r})\boldsymbol{\alpha}\cdot\nabla\psi_{\nu'}(\mathbf{r})=-\mathrm{i}c\hbar\int_{\mathbb{V}}\mathrm{d}^3\mathbf{r}\psi_{l\,\nu}^\dagger(\mathbf{r})\boldsymbol{\sigma}\cdot\nabla\psi_{s\,\nu'}(\mathbf{r})-\mathrm{i}c\hbar\int_{\mathbb{V}}\mathrm{d}^3\mathbf{r}\psi_{s\,\nu}^\dagger(\mathbf{r})\boldsymbol{\sigma}\cdot\nabla\psi_{l\,\nu'}(\mathbf{r}) \tag{A.40}$$

From the divergence theorem, for two functions $\phi(\mathbf{r})$ and $\chi(\mathbf{r})$,

$$\int_{\mathbb{V}}\mathrm{d}^3\mathbf{r}\,\nabla\left\{\phi^\dagger(\mathbf{r})\boldsymbol{\sigma}\chi(\mathbf{r})\right\}=\int_S\mathrm{d}^2\mathbf{r}_S\cdot\phi^\dagger(\mathbf{r}_S)\boldsymbol{\sigma}\chi(\mathbf{r}_S), \tag{A.41}$$

therefore

$$\int_{\mathbb{V}}\mathrm{d}^3\mathbf{r}\phi^\dagger(\mathbf{r})\boldsymbol{\sigma}\cdot\nabla\chi(\mathbf{r})=\int_S\mathrm{d}^2\mathbf{r}_S\cdot\phi^\dagger(\mathbf{r}_S)\boldsymbol{\sigma}\chi(\mathbf{r}_S)-\int_{\mathbb{V}}\mathrm{d}^3\mathbf{r}\left\{\nabla\phi(\mathbf{r})\right\}^\dagger\cdot\boldsymbol{\sigma}\chi(\mathbf{r}_S), \tag{A.42}$$

hence

$$\begin{aligned}\left[-\mathrm{i}c\hbar\boldsymbol{\alpha}\cdot\nabla\right]_{\nu\nu'}=&-\mathrm{i}c\hbar\int_S\mathrm{d}^2\mathbf{r}_S\psi_{l\,\nu}^\dagger(\mathbf{r}_S)\boldsymbol{\sigma}\psi_{s\,\nu'}(\mathbf{r}_S)\\&+\int_{\mathbb{V}}\mathrm{d}^3\mathbf{r}\left\{-\mathrm{i}c\hbar\boldsymbol{\sigma}\cdot\nabla\psi_{l\,\nu}(\mathbf{r})\right\}^\dagger\psi_{s\,\nu'}(\mathbf{r})\\&+\int_{\mathbb{V}}\mathrm{d}^3\mathbf{r}\psi_{s\,\nu}^\dagger(\mathbf{r})\left\{-\mathrm{i}c\hbar\boldsymbol{\sigma}\cdot\nabla\psi_{l\,\nu'}(\mathbf{r})\right\}.\end{aligned} \tag{A.43}$$

For plane waves

$$\begin{pmatrix}mc^2&-\mathrm{i}c\hbar\boldsymbol{\alpha}\cdot\nabla\\-\mathrm{i}c\hbar\boldsymbol{\alpha}\cdot\nabla&-mc^2\end{pmatrix}\begin{pmatrix}\psi_{l\,\nu}(\mathbf{r})\\\psi_{s\,\nu}(\mathbf{r})\end{pmatrix}=W_\nu\begin{pmatrix}\psi_{l\,\nu}(\mathbf{r})\\\psi_{s\,\nu}(\mathbf{r})\end{pmatrix} \tag{A.44}$$

where

$$W_\nu=\sqrt{c^2\hbar^2\left(|\mathbf{k}_\parallel|^2+|\mathbf{g}|^2+k_n^2\right)+\left(mc^2\right)^2}. \tag{A.45}$$

We can therefore put

$$-\mathrm{i}c\hbar\boldsymbol{\alpha}\cdot\nabla\psi_{s\,\nu}(\mathbf{r})=\left(W_\nu+mc^2\right)\psi_{l\,\nu}(\mathbf{r}),\quad-\mathrm{i}c\hbar\boldsymbol{\alpha}\cdot\nabla\psi_{l\,\nu}(\mathbf{r})=\left(W_\nu-mc^2\right)\psi_{s\,\nu}(\mathbf{r}). \tag{A.46}$$

Substituting this into (A.43) and adding (A.39), we have

$$\left[T\right]_{\nu\nu'}=\left(W_\nu+W_{\nu'}\right)\left[O\right]_{\nu\nu'}^{ss}+mc^2\left\{\left[O\right]_{\nu\nu'}^{ll}+\left[O\right]_{\nu\nu'}^{ss}\right\} \tag{A.47}$$

which is clearly Hermitian, and includes the surface contribution due to the discontinuity in the small component.

There is an important advantage in being able to express the interstitial Hamiltonian in terms of the overlap matrix. We do not have to calculate extra, complicated matrix elements in order to subtract the interstitial Hamiltonian from the atomic

spheres, we can just use the correction terms from the overlap matrix and substitute them into (A.47).

A.4.2 The Potential in the Interstitial

The interstitial potential is expanded in plane waves

$$V_{\text{int}}(\mathbf{r}) = \sum_{\mathbf{g}\,n} V_{\text{int}}(\mathbf{g}, n) \mathrm{e}^{\mathrm{i}\mathbf{g}\cdot\mathbf{r}_\parallel} \cos(k_n z) \tag{A.48}$$

and the matrix elements with the plane wave basis functions are

$$\begin{aligned}
\left[V_{\text{int}}\right]_{\nu\nu'} = \frac{2}{\mathcal{L}} \sum_{\mathbf{g}''} \sum_{n''} V_{\text{int}}(\mathbf{g}'', n'') \delta(\mathbf{g}'' - \mathbf{g} + \mathbf{g}') \\
\times \Bigg\{ \delta_{m_s m_s'} \int_{z_1}^{z_2} \mathrm{d}z \cos(k_{n''} z) \sin(k_n z) \sin(k_{n'} z) \\
+ \frac{1}{4} \gamma_k \gamma_{k'} \phi^\dagger_{m_s'} \boldsymbol{\sigma} \cdot \mathbf{k}^{(+)} \boldsymbol{\sigma} \cdot \mathbf{k}'^{(+)} \phi_{m_s} \int_{z_1}^{z_2} \mathrm{d}z \cos(k_{n''} z) \mathrm{e}^{-\mathrm{i}(k_n - k_{n'})z} \\
- \frac{1}{4} \gamma_k \gamma_k \phi^\dagger_{m_s} \boldsymbol{\sigma} \cdot \mathbf{k}^{(+)} \boldsymbol{\sigma} \cdot \mathbf{k}'^{(-)} \phi_{m_s'} \int_{z_1}^{z_2} \mathrm{d}z \cos(k_{n''} z) \mathrm{e}^{-\mathrm{i}(k_n + k_{n'})z} \\
- \frac{1}{4} \gamma_k \gamma_{k'} \phi^\dagger_{m_s} \boldsymbol{\sigma} \cdot \mathbf{k}^{(-)} \boldsymbol{\sigma} \cdot \mathbf{k}'^{(+)} \phi_{m_s'} \int_{z_1}^{z_2} \mathrm{d}z \cos(k_{n''} z) \mathrm{e}^{\mathrm{i}(k_n + k_{n'})z} \\
+ \frac{1}{4} \gamma_k \gamma_{k'} \phi^\dagger_{m_s} \boldsymbol{\sigma} \cdot \mathbf{k}^{(-)} \boldsymbol{\sigma} \cdot \mathbf{k}'^{(-)} \phi_{m_s'} \int_{z_1}^{z_2} \mathrm{d}z \cos(k_{n''} z) \mathrm{e}^{\mathrm{i}(k_n - k_{n'})z} \Bigg\}
\end{aligned} \tag{A.49}$$

In the surface part of the calculation, the interstitial potential is multiplied by a step function which is equal to 1 in the interstitial region and zero in the atomic spheres, as in standard LAPW calculations [137]. To do this, two expansions of the potential are made in the bulk LAPW calculation. The first is an expansion in plane waves throughout the whole interstitial region. The second is a spherical wave expansion in the atomic spheres. In the embedding potential calculation, the two separate expansions are used and the contribution from the spheres must be subtracted from the interstitial region.

A.4.3 The Spherical Part of the Hamiltonian in the Atomic Spheres

If we write the atomic solutions and their energy derivatives as

$$\chi_\Lambda(\mathbf{r}) = \begin{pmatrix} g_\kappa(r)\Omega_\Lambda(\hat{\mathbf{r}}) \\ \mathrm{i}f_\kappa(r)\Omega_{\overline{\Lambda}}(\hat{\mathbf{r}}) \end{pmatrix}, \quad \dot{\chi}_\Lambda(\mathbf{r}) = \begin{pmatrix} \dot{g}_\kappa(r)\Omega_\Lambda(\hat{\mathbf{r}}) \\ \mathrm{i}\dot{f}_\kappa(r)\Omega_{\overline{\Lambda}}(\hat{\mathbf{r}}) \end{pmatrix}, \quad \psi_\nu(\mathbf{r}) = \sum_{\Lambda\alpha} \{A^\alpha_{\Lambda\nu}\chi_\Lambda(\mathbf{r}) + B^\alpha_{\Lambda\nu}\dot{\chi}_\Lambda\} \tag{A.50}$$

we can use the fact that

$$\langle\chi_\Lambda|H|\chi_{\Lambda'}\rangle = W_{\kappa'}\langle\chi_\Lambda|\chi_{\Lambda'}\rangle + \mathrm{i}c\hbar\int_S \mathrm{d}^2\mathbf{r}_S \cdot \chi^\dagger_{l\,\Lambda}(\mathbf{r}_S)\boldsymbol{\sigma}\chi_{s\,\Lambda'}(\mathbf{r}_S) \tag{A.51}$$

where W_κ is the pivot energy at which the atomic basis functions are calculated and H includes the spherical part of the atomic potential. The integral is over the surface of the atomic sphere and comes from the fact that the small component can be discontinuous at S. We can then write

$$\begin{aligned}\frac{1}{2}\left\{\langle\chi_\Lambda|H|\chi_{\Lambda'}\rangle + \langle\chi_{\Lambda'}|H|\chi_\Lambda\rangle^\dagger\right\} &= \frac{1}{2}\left(W_\kappa + W_{\kappa'}\right)\langle\chi_\Lambda|\chi_{\Lambda'}\rangle \\ &\quad + \frac{\mathrm{i}c\hbar}{2}\int_S \mathrm{d}^2\mathbf{r}_S\left\{\chi^\dagger_{l\,\Lambda}(\mathbf{r}_S)\boldsymbol{\sigma}\chi_{s\,\Lambda'}(\mathbf{r}_S) - \chi^\dagger_{s\,\Lambda'}(\mathbf{r}_S)\boldsymbol{\sigma}\chi_{l\,\Lambda}(\mathbf{r}_S)\right\}\end{aligned} \tag{A.52}$$

and also note that

$$H|\dot{\chi}_\Lambda\rangle = W_\kappa|\dot{\chi}_\Lambda\rangle + |\chi_\Lambda\rangle. \tag{A.53}$$

Making use of these we have the matrix elements of the spherical part of the Hamiltonian in the atomic spheres:

$$[\,H_{sph}\,]_{\nu\nu'} = \sum_\alpha \left\{[A^\alpha_{\Lambda\nu}]^*\,A^\alpha_{\Lambda'\nu'}\,\mathcal{J} + [A^\alpha_{\Lambda\nu}]^*\,B^\alpha_{\Lambda'\nu'}\,\mathcal{K} + [B^\alpha_{\Lambda\nu}]^*\,A^\alpha_{\Lambda'\nu'}\,\mathcal{K} + [B^\alpha_{\Lambda\nu}]^*\,B^\alpha_{\Lambda'\nu'}\,\mathcal{L}\right\}, \tag{A.54}$$

where

$$\mathcal{J} = W_\kappa \int_0^{S_\alpha} \mathrm{d}r r^2 \left\{ g_\kappa^2(r) + f_\kappa^2(r) \right\} + c\hbar S_\alpha^2 g_\kappa(S_\alpha) f_\kappa(S_\alpha)$$

$$\mathcal{K} = W_\kappa \int_0^{S_\alpha} \mathrm{d}r r^2 \left\{ g_\kappa(r)\dot{g}_\kappa(r) + f_\kappa(r)\dot{f}_\kappa(r) \right\}$$

$$+ c\hbar S_\alpha^2 \left\{ \dot{g}_\kappa(S_\alpha) f_\kappa(S_\alpha) + g_\kappa(S_\alpha)\dot{f}_\kappa(S_\alpha) \right\}$$

$$+ \int_0^{S_\alpha} \mathrm{d}r r^2 \left\{ g_\kappa^2(r) + f_\kappa^2(r) \right\}$$

$$\mathcal{L} = W_\kappa \int_0^{S_\alpha} \mathrm{d}r r^2 \left\{ \dot{g}_\kappa^2(r) + \dot{f}_\kappa^2(r) \right\} + c\hbar S_\alpha^2 \dot{g}_\kappa(S_\alpha)\dot{f}_\kappa(S_\alpha)$$

$$+ \int_0^{S_\alpha} \mathrm{d}r r^2 \left\{ g_\kappa(r)\dot{g}_\kappa(r) + f_\kappa(r)\dot{f}_\kappa(r) \right\} \tag{A.55}$$

A.4.4 The Non-Spherical Potential in the Atomic Spheres

The non-spherical part of the potential in the α^{th} atomic sphere, $\Delta(\mathbf{r})$ is expanded in spherical harmonics

$$\Delta(\mathbf{r}) = \sum_\lambda \Delta_\lambda^\alpha(r) \sum_p Z_{\lambda p}^\alpha Y_{l_\lambda m_{\lambda p}}(\hat{\mathbf{r}}) \tag{A.56}$$

The matrix elements are

$$[\Delta]_{\nu\nu'} = \sum_\alpha \sum_{\Lambda\Lambda'} \sum_{\tau\tau'} \mathcal{A}_{\Lambda\nu\tau}^{\alpha *} \mathcal{A}_{\Lambda'\nu'\tau'}^{\alpha} \sum_\lambda \left\{ \Delta_{\tau\tau'}^{\alpha\kappa\kappa'\lambda} D_{\Lambda\Lambda'}^{\alpha\lambda} + \tilde{\Delta}_{\tau\tau'}^{\alpha\kappa\kappa'\lambda} D_{\overline{\Lambda\Lambda}'}^{\alpha\lambda} \right\} \tag{A.57}$$

Where we have put

$$\mathcal{A}_{\Lambda\nu\tau}^\alpha = \begin{cases} A_{\Lambda\nu\tau}^\alpha & \tau = 1 \\ B_{\Lambda\nu\tau}^\alpha & \tau = 2 \end{cases} . \tag{A.58}$$

The coefficients

$$\Delta_{\tau\tau'}^{\alpha\kappa\kappa'\lambda} = \int_0^{S_\alpha} \mathrm{d}r r^2 g_{\kappa\tau}(r) \Delta_\lambda^\alpha(r) g_{\kappa\tau'}(r) \tag{A.59}$$

and

$$\tilde{\Delta}_{\tau\tau'}^{\alpha\kappa\kappa'\lambda} = \int_0^{S_\alpha} \mathrm{d}r r^2 f_{\kappa\tau}(r) \Delta_\lambda^\alpha(r) f_{\kappa\tau'}(r) \tag{A.60}$$

where

$$g_{\kappa\tau}(r) = \begin{cases} g_\kappa(r) & \tau = 1 \\ \dot{g}_\kappa(r) & \tau = 2 \end{cases}, \qquad f_{\kappa\tau}(r) = \begin{cases} f_\kappa(r) & \tau = 1 \\ \dot{f}_\kappa(r) & \tau = 2 \end{cases} \tag{A.61}$$

The second set of coefficients are

$$D^{\alpha\lambda}_{\Lambda'\Lambda} = \sum_{m_s p} C_{\Lambda m_s} C_{\Lambda' m_s} Z^{\alpha}_{\lambda p} C^{l\mu - m_s}_{l'\mu' - m_s l_\lambda m_{\lambda p}}, \qquad D^{\alpha\lambda}_{\bar{\Lambda}'\bar{\Lambda}} = \sum_{m_s p} C_{\bar{\Lambda} m_s} C_{\bar{\Lambda}' m_s} Z^{\alpha}_{\lambda p} C^{\bar{l}\mu - m_s}_{\bar{l}'\mu' - m_s l_\lambda m_{\lambda p}} \tag{A.62}$$

where $C_{\Lambda\sigma}$ are Clebsch–Gordan coefficients and the Gaunt coefficients are given by

$$C^{l\mu - m_s}_{l'\mu' - m_s l_\lambda m_{\lambda p}} = \int_{4\pi} \mathrm{d}^2\hat{\mathbf{r}}\, Y^*_{l\mu - m_s}(\hat{\mathbf{r}}) Y_{l_\lambda m_{\lambda p}}(\hat{\mathbf{r}}) Y_{l'\mu' - m_s}(\hat{\mathbf{r}}). \tag{A.63}$$

A.5 Matrix Elements of the Embedding Potential

The embedding plane is chosen to be a plane so that the matrix elements are particularly easy to calculate, especially since the embedding potential is a 2×2 quantity, meaning that we only need to use the large component of the plane wave basis functions. In the relativistic LAPW basis they are given by

$$\left[\Gamma(W, \mathbf{k}_\parallel)\right]_{\nu\nu'} = \frac{2}{\mathbb{V}} \phi^\dagger_{m_s} \left[\Gamma(W, \mathbf{k}_\parallel)\right]_{\mu\mu'} \phi_{m'_s} \sin(k_n Z) \sin(k'_n Z) \tag{A.64}$$

where $\left[\Gamma(W, \mathbf{k}_\parallel)\right]_{\mu\mu'}$ is the matrix representation of $\Gamma(\mathbf{r}_S, \mathbf{r}'_S; W, \mathbf{k}_\parallel)$ indexed by $\mu = (\mathbf{g}, m_s)$, expanded on the plane at $z = Z$, given by (5.59). The sin dependence is replaced by cos for the cosine-like choice of plane waves.

Bibliography

[1] M. James and S. Crampin. Relativistic embedding method: The transfer matrix, complex band structures, transport, and surface calculations. *Phys. Rev. B*, 81(15):155439, Apr 2010.

[2] J E Inglesfield. A method of embedding. *Journal of Physics C: Solid State Physics*, 14(26):3795, 1981.

[3] S Crampin. An embedding scheme for the dirac equation. *Journal of Physics: Condensed Matter*, 16(49):8875, 2004.

[4] D. F. Lawden. *An Introduction to Tensor Calculus, Relativity and Cosmology.* John Wiley & Sons, 1982.

[5] J. P. Elliot & P. G. Dawber. *Symmetry in Physics.* The MacMillan Press Ltd, 1986.

[6] Walter Greiner and Stefan Schramm. Resource letter qedv-1: The qed vacuum. *American Journal of Physics*, 76(6):509–518, 2008.

[7] H. A. Lorentz. Electromagnetic phenomena in a system moving with any velocity less than that of light. *Proceedings of the Academy of Sciences of Amsterdam*, 6:809, 1904.

[8] S M McMurry. *Quantum Mechanics.* Addison Wesley Publishing Company, 1994.

[9] A Messiah. *Quantum Mechanics.* North Holland Publishing Company, fourth edition, 1970.

[10] P Strange. *Relativistic Quantum Mechanics.* Cambridge University Press, 1998.

[11] S. LaShell, B. A. McDougall, and E. Jensen. Spin splitting of an au(111) surface state band observed with angle resolved photoelectron spectroscopy. *Phys. Rev. Lett.*, 77(16):3419–3422, Oct 1996.

[12] D. Corson & F. Lorrain P. Lorrain. *Electromagnetic Fields and Waves.* W.H.Freeman and Company, 1988.

[13] Supriyo Datta and Biswajit Das. Electronic analog of the electro-optic modulator. *Applied Physics Letters*, 56(7):665–667, 1990.

[14] L. Petersen and P. Hedegard. A simple tight-binding model of spin-orbit splitting of sp-derived surface states. *Surface Science*, 459(1-2):49 – 56, 2000.

[15] J Henk, M Hoesch, J Osterwalder, A Ernst, and P Bruno. Spin–orbit coupling in the l-gap surface states of au(111): spin-resolved photoemission experiments and first-principles calculations. *Journal of Physics: Condensed Matter*, 16(43):7581, 2004.

[16] G. Nicolay, F. Reinert, S. Hüfner, and P. Blaha. Spin-orbit splitting of the l-gap surface state on au(111) and ag(111). *Phys. Rev. B*, 65(3):033407, Dec 2001.

[17] Friedrich Reinert. Spin–orbit interaction in the photoemission spectra of noble metal surface states. *Journal of Physics: Condensed Matter*, 15(5):S693, 2003.

[18] G. Bihlmayer, Yu.M. Koroteev, P.M. Echenique, E.V. Chulkov, and S. Blügel. The rashba-effect at metallic surfaces. *Surface Science*, 600(18):3888 – 3891, 2006. Berlin, Germany: 4-9 September 2005, Proceedings of the 23th European Conference on Surface Science.

[19] G. Binasch, P. Grünberg, F. Saurenbach, and W. Zinn. Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange. *Phys. Rev. B*, 39(7):4828–4830, Mar 1989.

[20] M. N. Baibich, J. M. Broto, A. Fert, F. Nguyen Van Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friederich, and J. Chazelas. Giant magnetoresistance of (001)fe/(001)cr magnetic superlattices. *Phys. Rev. Lett.*, 61(21):2472–2475, Nov 1988.

[21] S. A. Wolf, D. D. Awschalom, R. A. Buhrman, J. M. Daughton, S. von Molnar, M. L. Roukes, A. Y. Chtchelkanova, and D. M. Treger. Spintronics: A Spin-Based Electronics Vision for the Future. *Science*, 294(5546):1488–1495, 2001.

[22] Igor Zutc, Jaroslav Fabian, and S. Das Sarma. Spintronics: Fundamentals and applications. *Rev. Mod. Phys.*, 76(2):323–410, Apr 2004.

[23] J Hugo Dil. Spin and angle resolved photoemission on non-magnetic low-dimensional systems. *Journal of Physics: Condensed Matter*, 21(40):403001, 2009.

[24] Koichiro Yaji, Yoshiyuki Ohtsubo, Shinichiro Hatta, Hiroshi Okuyama, Koji Miyamoto, Taichi Okuda, Akio Kimura, Hirofumi Namatame, Masaki Taniguchi, and Tetsuya Aruga. Large rashba spin splitting of a metallic

surface-state band on a semiconductor surface. *Nat Commun*, 1(2):1–5, May 2010.

[25] Fabian Meier, Hugo Dil, Jorge Lobo-Checa, Luc Patthey, and Jürg Osterwalder. Quantitative vectorial spin analysis in angle-resolved photoemission: *bi?ag*(111) and *pb?ag*(111). *Phys. Rev. B*, 77(16):165431, Apr 2008.

[26] J. Hugo Dil, Fabian Meier, Jorge Lobo-Checa, Luc Patthey, Gustav Bihlmayer, and Jürg Osterwalder. Rashba-type spin-orbit splitting of quantum well states in ultrathin pb films. *Phys. Rev. Lett.*, 101(26):266802, Dec 2008.

[27] F Mandl and G Shaw. *Quantum Field Theory*. John Wiley & Sons, revised edition, 1993.

[28] E Corinaldesi and F Strocchi. *Relativistic Wave Mechanics*. North Holland Publishing Company, 1963.

[29] P. A. M. Dirac. The quantum theory of the electron. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 117(778):610–624, 1928.

[30] L I Schiff. *Quantum Mechanics*. McGraw Hill, 1968.

[31] W Greiner. *Relativistic Quantum Mechanics: Wave Equations*. Springer-Verlag, 3rd edition, 2000.

[32] T Takeda. Self-consistent relativistic bandstructure for gold. *Journal of Physics F: Metal Physics*, 10(6):1135, 1980.

[33] M. Abramowitz & I. A. Stegun. *Handbook of Mathematical Functions*. Dover Publications, 1972.

[34] G. B. Arfken & H. J. Weber. *Mathematical Methods for Physicists*. Harcourt academic Press, fifth edition, 2001.

[35] D D Koelling and B N Harmon. A technique for relativistic spin-polarised calculations. *Journal of Physics C: Solid State Physics*, 10(16):3107, 1977.

[36] E. Tamura. Relativistic single-site green function for general potentials. *Phys. Rev. B*, 45(7):3271–3278, Feb 1992.

[37] P A M Dirac. *The Principles of Quantum Mechanics*. Oxford University Press, 1958.

[38] E. Wigner. Über die operation der zeitumkehr in der quantenmechanik. *Nachrichten von der gesellschaft der wissenschaften zu Göttingen*, 1:546–559, 1932.

[39] G. Rickayzen. *Green's Functions and Condensed Matter*. Harcourt academic Press, 1980.

[40] S. Doniach & E.H. Sondheimer. *Green's Functions for Solid State Physicists*. Imperial College Press, 1998.

[41] P. T. Landsberg. *Solid State Theory: Methods and Applications*. Wiley, 1969.

[42] A. Gonis. *Green's Functions for Ordered and Disordered Systems*. North-Holland, 1992.

[43] J.E. Inglesfield. The electronic structure of surfaces with the matching green function method: I. general formalism. *Surface Science*, 76(2):355 – 378, 1978.

[44] T. Huhne, C. Zecha, H. Ebert, P. H. Dederichs, and R. Zeller. Full-potential spin-polarized relativistic korringa-kohn-rostoker method implemented and applied to bcc fe, fcc co, and fcc ni. *Phys. Rev. B*, 58(16):10236–10247, Oct 1998.

[45] R. M. Martin. *Electronic Structure*. Cambridge University Press, 2005.

[46] D. Wortmann, H. Ishida, and S. Blügel. Embedded green-function approach to the ballistic electron transport through an interface. *Phys. Rev. B*, 66(7):075113, Aug 2002.

[47] E. N. Economou. *Green's Functions in Quantum Physics*. Springer-Verlag, 3rd edition, 2006.

[48] K F Riley, M P Hobson, and S J Bence. *Mathematical Methods for Physics and Engineering*. Cambridge University Press, 1997.

[49] M. Born & R. Oppenheimer. On the quantum theory of molecules. *Annalen der Physik*, 84:458–490, 1927.

[50] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136(3B):B864–B871, Nov 1964.

[51] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133–A1138, Nov 1965.

[52] Donald G. Anderson. Iterative procedures for nonlinear integral equations. *J. ACM*, 12(4):547–560, 1965.

[53] C. G. Broyden. A class of methods for solving nonlinear simultaneous equations. *Mathematics of Computation*, 19(92):577–593, 1965.

[54] D. D. Johnson. Modified broyden's method for accelerating convergence in self-consistent calculations. *Phys. Rev. B*, 38(18):12807–12813, Dec 1988.

[55] J. C. Slater. A simplification of the hartree-fock method. *Phys. Rev.*, 81(3):385–390, Feb 1951.

[56] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.*, 45(7):566–569, Aug 1980.

[57] B. Ceperley, D. & Alder. Quantum monte carlo. *Science*, 231(4738):555–560, 1986.

[58] J. P. Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23(10):5048–5079, May 1981.

[59] John P. Perdew and Yue Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B*, 45(23):13244–13249, Jun 1992.

[60] L. Wilkes & N. Nusair S. H. Vosko. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadin Journal of Physics*, 58:1200–1207, 1980.

[61] R. O. Jones and O. Gunnarsson. The density functional formalism, its applications and prospects. *Rev. Mod. Phys.*, 61(3):689–746, Jul 1989.

[62] John P. Perdew and Wang Yue. Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation. *Phys. Rev. B*, 33(12):8800–8802, Jun 1986.

[63] L. J. Sham, Z. Reut, J. W. Wilkins, V. Heine, L. M. Falicov, and D. M. Edwards. Density functionals beyond the local approximation [and discussion]. *Philosophical Transactions: Physical Sciences and Engineering*, 334(1635):481–490, 1991.

[64] R. W. Godby, M. Schlüter, and L. J. Sham. Self-energy operators and exchange-correlation potentials in semiconductors. *Phys. Rev. B*, 37(17):10159–10175, Jun 1988.

[65] James A. Snyder, Dominic R. Alfonso, John E. Jaffe, Zijing Lin, Anthony C. Hess, and Maciej Gutowski. Periodic density functional lda and gga study of co adsorption at the (001) surface of mgo. *The Journal of Physical Chemistry B*, 104(19):4717–4722, 2000.

[66] A. Svane and O. Gunnarsson. Transition-metal oxides in the self-interaction–corrected density-functional formalism. *Phys. Rev. Lett.*, 65(9):1148–1151, Aug 1990.

[67] F. Aryasetiawan and O. Gunnarsson. Electronic structure of nio in the *gw* approximation. *Phys. Rev. Lett.*, 74(16):3221–3224, Apr 1995.

[68] John P. Perdew. Accurate density functional for the energy: Real-space cutoff of the gradient expansion for the exchange hole. *Phys. Rev. Lett.*, 55(16):1665–1668, Oct 1985.

[69] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77(18):3865–3868, Oct 1996.

[70] John P. Perdew, J. A. Chevary, S. H. Vosko, Koblar A. Jackson, Mark R. Pederson, D. J. Singh, and Carlos Fiolhais. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B*, 46(11):6671–6687, Sep 1992.

[71] F Aryasetiawan and O Gunnarsson. The gw method. *Reports on Progress in Physics*, 61(3):237, 1998.

[72] A. K. Rajagopal and J. Callaway. Inhomogeneous electron gas. *Phys. Rev. B*, 7(5):1912–1919, Mar 1973.

[73] A K Rajagopal. Inhomogeneous relativistic electron gas. *Journal of Physics C: Solid State Physics*, 11(24):L943, 1978.

[74] A H MacDonald and S H Vosko. A relativistic density functional formalism. *Journal of Physics C: Solid State Physics*, 12(15):2977, 1979.

[75] A. H. MacDonald, J. M. Daams, S. H. Vosko, and D. D. Koelling. Non-muffin-tin and relativistic interaction effects on the electronic structure of noble metals. *Phys. Rev. B*, 25(2):713–725, Jan 1982.

[76] A. Facco Bonetti H. Müller R. M. Dreizler E. Engel, S. Keller. Local and nonlocal relativistic exchange-correlation energy functionals: Comparison to relativistic optimized-potential-model results. *Physical Review A*, 52(4):2750–2764, 1995.

[77] E. Engel and R. Dreizler. Relativistic density functional theory. In R. Nalewajski, editor, *Density Functional Theory II*, volume 181 of *Topics in Current Chemistry*, pages 1–80. Springer Berlin / Heidelberg, 1996. 10.1007/BFb0016642.

[78] Michael Springborg. *Methods of Electronic Structure Calculations*. John Wiley & Sons, 2000.

[79] Karl D. Brommer, M. Needels, B. Larson, and J. D. Joannopoulos. Ab initio theory of the si(111)-(7×7) surface reconstruction: A challenge for massively parallel computation. *Phys. Rev. Lett.*, 68(9):1355–1358, Mar 1992.

[80] G. Brocks, P. J. Kelly, and R. Car. Binding and diffusion of a si adatom on the si(100) surface. *Phys. Rev. Lett.*, 66(13):1729–1732, Apr 1991.

[81] P Stoltze. Simulation of surface defects. *Journal of Physics: Condensed Matter*, 6(45):9495, 1994.

[82] H. L. Skriver and N. M. Rosengaard. Surface energy and work function of elemental metals. *Phys. Rev. B*, 46(11):7157–7168, Sep 1992.

[83] F. E. Olsson, M. Persson, A. G. Borisov, J.-P. Gauyacq, J. Lagoute, and S. Fölsch. Localization of the cu(111) surface state by single cu adatoms. *Phys. Rev. Lett.*, 93(20):206803, Nov 2004.

[84] R. A. Evarestov, S. Piskunov, E. A. Kotomin, and G. Borstel. Single impurities in insulators: Ab initio study of fe-doped $srtio_3$. *Phys. Rev. B*, 67(6):064101, Feb 2003.

[85] M.J. Mehl and B.M. Klein. All-electron first-principles supercell total-energy calculation of the vacancy formation energy in aluminium. *Physica B: Condensed Matter*, 172(1-2):211 – 215, 1991.

[86] N. Shanthi and D. D. Sarma. Electronic structure of electron doped $srtio_3$: $srtio_{3-\delta}$ and $sr_{1-x}la_xtio_3$. *Phys. Rev. B*, 57(4):2153–2158, Jan 1998.

[87] J. D. Burton, S. S. Jaswal, E. Y. Tsymbal, O. N. Mryasov, and O. G. Heinonen. Atomic and electronic structure of the cofeb/mgo interface from first principles. *Applied Physics Letters*, 89(14):142507, 2006.

[88] V. Drchal, J. Kudrnovský, P. Bruno, P. H. Dederichs, I. Turek, and P. Weinberger. Electron transport in magnetic multilayers: Effect of disorder. *Phys. Rev. B*, 65(21):214414, May 2002.

[89] J. E. Inglesfield. Embedding at surfaces. *Computer Physics Communications*, 137(1):89 – 107, 2001.

[90] F. Forster, A. Bendounan, F. Reinert, V.G. Grigoryan, and M. Springborg. The shockley-type surface state on ar covered au(111): High resolution photoemission results and the description by slab-layer dft calculations. *Surface Science*, 601(23):5595 – 5604, 2007.

[91] J. Korringa. On the calculation of the energy of a bloch wave in a metal. *Physica*, 13(6-7):392 – 400, 1947.

[92] W. Kohn and N. Rostoker. Solution of the schrödinger equation in periodic lattices with an application to metallic lithium. *Phys. Rev.*, 94(5):1111–1120, Jun 1954.

[93] Phivos Mavropoulos, Nikolaos Papanikolaou, and Peter H. Dederichs. Korringa-kohn-rostoker green-function formalism for ballistic transport. *Phys. Rev. B*, 69(12):125104, Mar 2004.

[94] B. Yu. Yavorsky and I. Mertig. Noncollinear interface magnetism and ballistic transport in fe/feo/mgo/fe tunnel junctions: Ab initio calculations using the kkr method. *Phys. Rev. B*, 74(17):174402, Nov 2006.

[95] W. H. Butler. Theory of electronic transport in random alloys: Korringa-kohn-rostoker coherent-potential approximation. *Phys. Rev. B*, 31(6):3260–3277, Mar 1985.

[96] N Papanikolaou, R Zeller, and P H Dederichs. Conceptual improvements of the kkr method. *Journal of Physics: Condensed Matter*, 14(11):2799, 2002.

[97] N. Papanikolaou, R. Zeller, P. H. Dederichs, and N. Stefanou. Lattice distortion in cu-based dilute alloys: A first-principles study by the kkr green-function method. *Phys. Rev. B*, 55(7):4157–4167, Feb 1997.

[98] L. Szunyogh, B. Újfalussy, P. Weinberger, and J. Kollár. Self-consistent localized kkr scheme for surfaces and interfaces. *Phys. Rev. B*, 49(4):2721–2729, Jan 1994.

[99] J. M. MacLaren, S. Crampin, D. D. Vvedensky, and J. B. Pendry. Layer korringa-kohn-rostoker technique for surface and interface electronic properties. *Phys. Rev. B*, 40(18):12164–12175, Dec 1989.

[100] L. Diekhöner, M. A. Schneider, A. N. Baranov, V. S. Stepanyuk, P. Bruno, and K. Kern. Surface states of cobalt nanoislands on cu(111). *Phys. Rev. Lett.*, 90(23):236801, Jun 2003.

[101] Samir Lounis, Phivos Mavropoulos, Peter H. Dederichs, and Stefan Blügel. Surface-state scattering by adatoms on noble metals: Ab initio calculations using the korringa-kohn-rostoker green function method. *Phys. Rev. B*, 73(19):195421, May 2006.

[102] M. Asato, A. Settels, T. Hoshino, T. Asada, S. Blügel, R. Zeller, and P. H. Dederichs. Full-potential kkr calculations for metals and semiconductors. *Phys. Rev. B*, 60(8):5202–5210, Aug 1999.

[103] Miki Nagano, Ayaka Kodama, T Shishidou, and T Oguchi. A first-principles study on the rashba effect in surface systems. *Journal of Physics: Condensed Matter*, 21(6):064239, 2009.

[104] G A Benesh and J E Inglesfield. An embedding approach for surface calculations. *Journal of Physics C: Solid State Physics*, 17(9):1595, 1984.

[105] S Crampin, J B A N van Hoof, M Nekovee, and J E Inglesfield. Full-potential embedding for surfaces and interfaces. *Journal of Physics: Condensed Matter*, 4(6):1475, 1992.

[106] J. E. Inglesfield and G. A. Benesh. Surface electronic structure: Embedded self-consistent calculations. *Phys. Rev. B*, 37(12):6682–6700, Apr 1988.

[107] J. E. Inglesfield and G. A. Benesh. Bulk and surface states on al(001). *Surface Science*, 200(2-3):135 – 143, 1988.

[108] G.A. Benesh and L.S.G. Liyanage. The surface electronic structure of oxygen on pt(001)(1x1). *Surface Science*, 261(1-3):207 – 216, 1992.

[109] H. Ishida. Surface-embedded green function calculation using non-local pseudopotentials. *Surface Science*, 388(1-3):71 – 83, 1997.

[110] H. Ishida. Surface-embedded green-function method: A formulation using a linearized-augmented-plane-wave basis set. *Phys. Rev. B*, 63(16):165409, Apr 2001.

[111] G A Benesh, L S G Liyanage, and J C Pingel. The surface electronic structure of (1*1) pt(001). *Journal of Physics: Condensed Matter*, 2(46):9065, 1990.

[112] G. A. Benesh and L. S. G. Liyanage. Surface-embedded green-function method for general surfaces: Application to al(111). *Phys. Rev. B*, 49(24):17264–17272, Jun 1994.

[113] M. Nekovee, S. Crampin, and J. E. Inglesfield. Magnetic splitting of image states at fe(110). *Phys. Rev. Lett.*, 70(20):3099–3102, May 1993.

[114] G.C. Aers and J.E. Inglesfield. Electric field and ag(001) surface electronic structure. *Surface Science*, 217(1-2):367 – 383, 1989.

[115] H. Ishida, A. V. Petukhov, and A. Liebsch. Surface anisotropy in optical second harmonic generation ii. embedding approach to al(111) and vicinal al(001) surfaces. *Surface Science*, 340(1-2):1 – 15, 1995.

[116] H. Ishida and A. Liebsch. Electronic excitations in alkali-metal overlayers. i. unreconstructed low-temperature phase of li/al. *Phys. Rev. B*, 57(19):12550–12557, May 1998.

[117] J. E. Inglesfield, S. Crampin, and H. Ishida. Embedding potential definition of channel functions. *Phys. Rev. B*, 71(15):155120, Apr 2005.

[118] D. Wortmann, H. Ishida, and S. Blügel. Ab initio green-function formulation of the transfer matrix: Application to complex band structures. *Phys. Rev. B*, 65(16):165103, Mar 2002.

[119] O. R. Davies and J. E. Inglesfield. Embedding method for conductance of dna. *Phys. Rev. B*, 69(19):195110, May 2004.

[120] Antonis N. Andriotis and Madhu Menon. Green's function embedding approach to quantum conductivity of single wall carbon nanotubes. *The Journal of Chemical Physics*, 115(6):2737–2742, 2001.

[121] S. Crampin, M. Nekovee, and J. E. Inglesfield. Embedding method for confined quantum systems. *Phys. Rev. B*, 51(11):7318–7320, Mar 1995.

[122] Jiutao Li, Wolf-Dieter Schneider, Richard Berndt, and S. Crampin. Electron confinement to nanoscale ag islands on ag(111): A quantitative study. *Phys. Rev. Lett.*, 80(15):3332–3335, Apr 1998.

[123] M. I. Trioni, G. P. Brivio, S. Crampin, and J. E. Inglesfield. Embedding approach to the isolated adsorbate. *Phys. Rev. B*, 53(12):8052–8064, Mar 1996.

[124] G. A. Benesh and Daniel Gebreselasie. Relaxation of al(001) and al(110): Surface embedded green function total-energy and force calculation. *Phys. Rev. B*, 54(8):5940–5945, Aug 1996.

[125] Holger Wallmeier and Werner Kutzelnigg. Basis-set expansion of the dirac equation without variational collapse: Numerical test of the forth-back free-particle foldy-wouthuysen transformation. *Phys. Rev. A*, 28(5):3092–3094, Nov 1983.

[126] Werner Kutzelnigg. Basis set expansion of the dirac operator without variational collapse. *International Journal of Quantum Chemistry*, 25:107–129, 1984.

[127] Richard E. Stanton and Stephen Havriliak. Kinetic balance: A partial solution to the problem of variational safety in dirac calculations. *The Journal of Chemical Physics*, 81(4):1910–1918, 1984.

[128] H. Schwarz, W. H. E. & Wallmeier. Basis set expansions of relativistic molecular wave equations. *Molecular Physics*, 46(5):1045, 1982.

[129] G. W. F. Drake and S. P. Goldman. Application of discrete-basis-set methods to the dirac equation. *Phys. Rev. A*, 23(5):2093–2098, May 1981.

[130] S. P. Goldman. Variational representation of the dirac-coulomb hamiltonian with no spurious roots. *Phys. Rev. A*, 31(6):3541–3549, Jun 1985.

[131] Ian P. Grant. Conditions for convergence of variational solutions of dirac's equation in a finite basis. *Phys. Rev. A*, 25(2):1230–1232, Feb 1982.

[132] H. Weyl. *The Theory of Groups and Quantum Mechanics*. Dover Publications, 1931.

[133] D Hegarty and P J C Aerts. Variational solutions of the dirac equation for atoms and molecules. *Physica Scripta*, 36(3):432, 1987.

[134] A. McD. Mercer and Peter R. Mercer. Cauchy's interlace theorem and lower bounds for the spectral radius. *International Journal of Mathematics and Mathematical Sciences*, 23(8):563–566, 2000.

[135] J. K. L. MacDonald. Successive approximations by the rayleigh-ritz variation method. *Phys. Rev.*, 43(10):830–833, May 1933.

[136] Edward S. Sachs, Juergen Hinze, and Nora H. Sabelli. Frozen core approximation, a pseudopotential method tested on six states of nah. *The Journal of Chemical Physics*, 62(9):3393–3398, 1975.

[137] D. J. Singh and L. Nordström. *Planewaves, Pseudopotentials and the LAPW method*. Springer Verlag, 2nd edition, 2006.

[138] D. R. Hamann, M. Schlüter, and C. Chiang. Norm-conserving pseudopotentials. *Phys. Rev. Lett.*, 43(20):1494–1497, Nov 1979.

[139] Nicholas J. Ramer and Andrew M. Rappe. Designed nonlocal pseudopotentials for enhanced transferability. *Phys. Rev. B*, 59(19):12471–12478, May 1999.

[140] J. C. Slater. An augmented plane wave method for the periodic potential problem. *Phys. Rev.*, 92(3):603–608, Nov 1953.

[141] J. C. Slater. Wave functions in a periodic potential. *Phys. Rev.*, 51(10):846–851, May 1937.

[142] O. Krogh Andersen. Linear methods in band theory. *Phys. Rev. B*, 12(8):3060–3083, Oct 1975.

[143] T. Ohwaki, H. Ishida, and A. Liebsch. First-principles calculation of field emission from metal surfaces. *Phys. Rev. B*, 68(15):155422, Oct 2003.

[144] T Takeda. Linear methods for fully relativistic energy-band calculations. *Journal of Physics F: Metal Physics*, 9(5):815, 1979.

[145] C. Jirauschek. Accuracy of transfer matrix approaches for solving the effective mass schrodinger equation. *Quantum Electronics, IEEE Journal of*, 45(9):1059–1067, sep. 2009.

[146] Yia-Chung Chang. Complex band structures of zinc-blende materials. *Phys. Rev. B*, 25(2):605–619, Jan 1982.

[147] V Heine. On the general theory of surface states and scattering of electrons in solids. *Proceedings of the Physical Society*, 81(2):300, 1963.

[148] Timothy B. Boykin. Generalized eigenproblem method for surface and interface states: The complex bands of gaas and alas. *Phys. Rev. B*, 54(11):8107–8115, Sep 1996.

[149] Ph. Mavropoulos, N. Papanikolaou, and P. H. Dederichs. Complex band structure and tunneling through ferromagnet /insulator /ferromagnet junctions. *Phys. Rev. Lett.*, 85(5):1088–1091, Jul 2000.

[150] John K. Tomfohr and Otto F. Sankey. Complex band structure, decay lengths, and fermi level alignment in simple molecular electronic systems. *Phys. Rev. B*, 65(24):245105, May 2002.

[151] Tongsheng Xia, L.F. Register, and S.K. Banerjee. Quantum transport in double-gate mosfets with complex band structure. *Electron Devices, IEEE Transactions on*, 50(6):1511 – 1516, jun. 2003.

[152] Pawel Pomorski, Christopher Roland, and Hong Guo. Quantum transport through short semiconducting nanotubes: A complex band structure analysis. *Phys. Rev. B*, 70(11):115408, Sep 2004.

[153] José M. Soler and Arthur R. Williams. Simple formula for the atomic forces in the augmented-plane-wave method. *Phys. Rev. B*, 40(3):1560–1564, Jul 1989.

[154] N. Egede Christensen and B. O. Seraphin. Relativistic band calculation and the optical properties of gold. *Phys. Rev. B*, 4(10):3321–3344, Nov 1971.

[155] D. K. Ferry & S. K. Goodnick. *Transport in Nanostructures*. Cambridge University Press, 2001.

[156] Ryogo Kubo. Statistical-mechanical theory of irreversible processes. i. general theory and simple applications to magnetic and conduction problems. *Journal of the Physical Society of Japan*, 12(6):570–586, 1957.

[157] P. Weinberger, J. Zabloudil, R. H. Hammerling, L. Szunyogh, T. L. Monchesky, and B. Heinrich. Role of interfaces in the magnetoresistance of au/fe/au/fe multilayers. *Phys. Rev. B*, 67(5):054404, Feb 2003.

[158] C. Blaas, P. Weinberger, L. Szunyogh, P. M. Levy, and C. B. Sommers. Ab initio calculations of magnetotransport for magnetic multilayers. *Phys. Rev. B*, 60(1):492–501, Jul 1999.

[159] C. Blaas, L. Szunyogh, P. Weinberger, C. Sommers, P. M. Levy, and J. Shi. Theoretical evaluation of magnetotransport properties in co/cu/co-based spin valves. *Phys. Rev. B*, 65(13):134427, Mar 2002.

[160] F. Rao and A.J. Freeman. Gmr in magnetic multilayers from a first principles band structure kubo-greenwood approach. *Magnetics, IEEE Transactions on*, 34(4):930 –932, jul 1998.

[161] Kees M. Schep, Paul J. Kelly, and Gerrit E. W. Bauer. Ballistic transport and electronic structure. *Phys. Rev. B*, 57(15):8907–8926, Apr 1998.

[162] R. Landauer. Spatial variation of currents and fields due to localized scatterers in metallic conduction. *IBM Journal of Research and Development*, 1(3):223 –231, july 1957.

[163] M. Büttiker. Four-terminal phase-coherent conductance. *Phys. Rev. Lett.*, 57(14):1761–1764, Oct 1986.

[164] David C. Langreth and Elihu Abrahams. Derivation of the landauer conductance formula. *Phys. Rev. B*, 24(6):2978–2984, Sep 1981.

[165] K. Xia, M. Zwierzycki, M. Talanana, P. J. Kelly, and G. E. W. Bauer. First-principles scattering matrices for spin transport. *Phys. Rev. B*, 73(6):064420, Feb 2006.

[166] P. X. Xu, V. M. Karpan, K. Xia, M. Zwierzycki, I. Marushchenko, and P. J. Kelly. Influence of roughness and disorder on tunneling magnetoresistance. *Phys. Rev. B*, 73(18):180402, May 2006.

[167] T. P. Pareek and A. M. Jayannavar. Generation and measurement of nonequilibrium spin currents in two-terminal systems. *Phys. Rev. B*, 77(15):153307, Apr 2008.

[168] M. Zaffalon, Aveek Bid, M. Heiblum, D. Mahalu, and V. Umansky. Transmission phase of a singly occupied quantum dot in the kondo regime. *Phys. Rev. Lett.*, 100(22):226601, Jun 2008.

[169] D Wortmann, G Bihlmayer, and S Blügel. Ab initio calculations of interface effects in tunnelling through mgo barriers on fe(100). *Journal of Physics: Condensed Matter*, 16(48):S5819, 2004.

[170] H. Ishida, D. Wortmann, and T. Ohwaki. First-principles calculations of tunneling conductance. *Phys. Rev. B*, 70(8):085409, Aug 2004.

[171] H. Ishida. Embedded green-function calculation of the conductance of oxygen-incorporated au and ag monatomic wires. *Phys. Rev. B*, 75(20):205419, May 2007.

[172] J. M. MacLaren, X.-G. Zhang, W. H. Butler, and Xindong Wang. Layer kkr approach to bloch-wave transmission and reflection: Application to spin-dependent tunneling. *Phys. Rev. B*, 59(8):5470–5478, Feb 1999.

[173] Gerhard Wachutka. General solution theory for schrödinger's equation in arbitrary 2d-periodic spatial structures ii. the synthesis of global solutions by layer composition. *Annals of Physics*, 187(2):314 – 354, 1988.

[174] H. Bross. The final states in photoemission and field emission. *Zeitschrift für Physik B Condensed Matter*, 28:173–182, 1977. 10.1007/BF01313039.

[175] William Shockley. On the surface states associated with a periodic potential. *Phys. Rev.*, 56(4):317–323, Aug 1939.

[176] J E Inglesfield. Surface electronic structure. *Reports on Progress in Physics*, 45(3):223, 1982.

[177] S. L. Cunningham. Special points in the two-dimensional brillouin zone. *Phys. Rev. B*, 10(12):4988–4994, Dec 1974.

[178] M. Weinert. Solution of poisson's equation: Beyond ewald-type methods. *Journal of Mathematical Physics*, 22(11):2433–2439, 1981.

[179] Riccardo Mazzarello, Andrea Dal Corso, and Erio Tosatti. Spin-orbit modifications and splittings of deep surface states on clean au(111). *Surface Science*, 602(4):893 – 905, 2008.

[180] T. P. Pareek and P. Bruno. Spin coherence in a two-dimensional electron gas with rashba spin-orbit interaction. *Phys. Rev. B*, 65(24):241305, May 2002.

[181] John Schliemann, J. Carlos Egues, and Daniel Loss. Nonballistic spin-field-effect transistor. *Phys. Rev. Lett.*, 90(14):146801, Apr 2003.

[182] Matthias Muntwiler, Moritz Hoesch, Vladimir N. Petrov, Matthias Hengsberger, Luc Patthey, Ming Shi, Mihaela Falub, Thomas Greber, and Jürg Osterwalder. Spin- and angle-resolved photoemission spectroscopy study of the au(111) shockley surface state. *Journal of Electron Spectroscopy and Related Phenomena*, 137-140:119 – 123, 2004. ICESS-9 Proceedings of the 9th International Conference on Electronic Spectroscopy and Structure.

[183] M. Hoesch, M. Muntwiler, V. N. Petrov, M. Hengsberger, L. Patthey, M. Shi, M. Falub, T. Greber, and J. Osterwalder. Spin structure of the shockley surface state on $au(111)$. *Phys. Rev. B*, 69(24):241401, Jun 2004.

[184] J R McLaughlan, E M Llewellyn-Samuel, and S Crampin. Spin–orbit splitting of image states. *Journal of Physics: Condensed Matter*, 16(39):6841, 2004.

[185] J. Henk, A. Ernst, and P. Bruno. Spin polarization of the l-gap surface states on au(111). *Phys. Rev. B*, 68(16):165416, Oct 2003.

[186] Hendrik J. Monkhorst and James D. Pack. Special points for brillouin-zone integrations. *Phys. Rev. B*, 13(12):5188–5192, Jun 1976.

[187] A. Mugarza, A. Mascaraque, V. Repain, S. Rousset, K. N. Altmann, F. J. Himpsel, Yu. M. Koroteev, E. V. Chulkov, F. J. García de Abajo, and J. E. Ortega. Lateral quantum wells at vicinal au(111) studied with angle-resolved photoemission. *Phys. Rev. B*, 66(24):245419, Dec 2002.

[188] Tamio Oguchi and Tatsuya Shishidou. The surface rashba effect: a k · p perturbation approach. *Journal of Physics: Condensed Matter*, 21(9):092001, 2009.

[189] T. Eguchi, A. Kamoshida, M. Ono, M. Hamada, R. Shoda, T. Nishio, A. Harasawa, T. Okuda, T. Kinoshita, and Y. Hasegawa. Surface states of a pd monolayer formed on a au(111) surface studied by angle-resolved photoemission spectroscopy. *Phys. Rev. B*, 74(7):073403, Aug 2006.

[190] Yu. M. Koroteev, G. Bihlmayer, J. E. Gayone, E. V. Chulkov, S. Blügel, P. M. Echenique, and Ph. Hofmann. Strong spin-orbit splitting on bi surfaces. *Phys. Rev. Lett.*, 93(4):046403, Jul 2004.

THANKS

First of all, thanks go to my supervisor, Simon Crampin for all his hard work and patience. I would also like to thank Prof. Hiroshi Ishida for providing the code that I made a right mess of and for explaining what some bits of it were for. Thanks to office colleagues for help with computers and things. Thanks also to friends and family for not forgetting I existed during my hermit like status while writing this thesis. Last but not least, I am eternally grateful to Charlotte for not having any interest in the embedding method, and for the amazing support she gave despite this.